AD-A103 955 PITTSBURGH UNIV PA INST FOR COMPUTATIONAL MATHEMATICS--ETC F/G 13/13
COMPUTATIONAL ERROR ESTIMATES AND ADAPTIVE PROCESSES FOR SOME N--ETC(U)
JUL 81 I BABUSKA, W C RHEINBOLDT N00014-80-C-0455
UNCLASSIFIED ICMA-81-28 NL

1 OF 1
AD A
103955

END
DATE
FILMED
10-81
DTIC

LEVEL II (12)

AD A103955

INSTITUTE FOR COMPUTATIONAL MATHEMATICS AND APPLICATIONS

Department of Mathematics and Statistics

University of Pittsburgh

DTIC
ELECTE
SEP 9 1981
S D
D

81 8 05 80

DTIC FILE COPY

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

Accession For
NTIS GRA&I ☒
DTIC TAB ☐
Unannounced ☐
Justification
By Per Ltr. on file
Distribution/
Availability Codes
Dist | Avail and/or Special
A

Technical Report, ICMA-81-28

COMPUTATIONAL ERROR ESTIMATES AND ADAPTIVE PROCESSES FOR SOME NONLINEAR STRUCTURAL PROBLEMS*

Ivo Babuska
Institute for Physical Science and Technology
University of Maryland, College Park, MD 20742

and

Werner C. Rheinboldt
Department of Mathematics and Statistics
University of Pittsburgh, Pittsburgh, PA 15260

July 1981

*This work was in part supported under ONR contracts N0014-77-C-0623 and N0014-80-C-0455 and NSF grant MCS-78-05299.

DTIC
ELECTE
SEP 9 1981
S D
D

Approved for public release;
Distribution Unlimited

Abstract

In earlier papers the authors introduced and analyzed the calculation of reliable, a posteriori error estimates for finite element solutions and discussed the design and effectivity of adaptive procedures based upon them. While most of this work concerned linear problems, this paper is intended to show that the same approaches remain also highly effective in the nonlinear case. As a model problem a planar, elastic rod is considered involving both material as well as geometrical nonlinearities; but, as far as possible the discussion is kept general. For nonlinear problems of this type interest centers on an analysis of the shape and features of the equilibrium surface. After some general remarks about such surfaces and the related computational problems, a general continuation process is summarized and some of its extensions for determining limit points and tracing stability paths are discussed. Then estimators are introduced for the error along the solution paths and of the computed critical values. Experimental results for the model problem show the effectivity of these estimators and of the adaptive procedures based upon them. Then some aspects relating to the occurrence and identification of spurious solutions are discussed. In order to illustrate desirable generalizations of these results, the paper ends with an outline of some results for linear problems of the type that should be achievable also in the nonlinear case.

Computational Error Estimates and Adaptive Processes for Some Nonlinear Structural Problems[1)]

by

Ivo Babuska[2)] and Werner C. Rheinboldt[3)]

1. Introduction

In general, the objective of a computational solution of a physical problem is to obtain an acceptably accurate and reliable prediction of the behavior of the physical phenomena under study. For this the computation should produce not only an approximate numerical solution but also some reliable information about its accuracy. More specifically, this accuracy should be assessed in comparison with the solution of a reasonably general mathematical model of the problem and in terms of a specified error norm that matches the requirements of the study.

The effectivity of such an estimation may be judged by the effectivity index

$$\theta = \frac{\text{estimated error}}{\text{true error}} . \tag{1.1}$$

In practice it is usually more important for θ to be close to one than that

---

1) This work was in part supported under ONR contracts N0014-77-C-0623 and N0014-80-C-0455 and NSF grant MCS-78-05299.

2) Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742.

3) Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA 15260.

$\theta \geq 1$. Moreover, it is essential that θ converges to one when the errors converge to zero, so that for an accuracy of say 5% or 10% the value of $|\theta-1|$ is expected to be less than, say, 0.1 or 0.2. The requirement of designing error estimators with these properties for realistic classes of problems and various different norms certainly represents a demanding task.

In order to achieve the stated objective of the computation adaptive processes appear to be optimal tools. In fact, it is beginning to be a widely accepted observation that for realistic problems it is rarely feasible to design numerical processes which reliably and effectively achieve the desired accuracy in the prescribed norm and yet which do not utilize some form of adaptivity. Unfortunately, the term "adaptation" remains to be rather ill-defined and is often used very loosely. A clarification of this concept may require close attention to results in such fields as automatic control theory, artificial intelligence, learning processes, etc. (see eg. [22], [33], [42], [44]). In our context, the availability of precisely defined, reliable error estimators appears to be central to the design of effective adaptive procedures. The effectivity has to be measured here in terms of the size of the class of problems for which solutions within a specified range of accuracy are obtained with minimal cost.

These general concepts for the numerical solution of certain practical problems have been discussed by us in several recent papers. More specifically, for the finite element solution of certain problems in structural mechanics we introduced and analyzed the calculation of reliable, a posteriori error estimates, [4], [5], [6], [9], [13], as well as the design and effectivity of adaptive procedures based on these estimates, [7], [8], [10], [11], [12], [46].

Up to now essentially all this work concerned linear problems in one or two space dimension. Not unexpectedly, for nonlinear problems the

situation is considerably more difficult and much remains to be done. In the second part of [12] some results were given for the finite element solution of a particular nonlinear model problem in R^1 which showed that a posteriori error estimates and adaptive approaches are also highly effective in such nonlinear cases even though the corresponding theory is as yet far from complete.

This paper is intended to be a continuation of this work on nonlinear problems. Such nonlinear problems show special features not present in the linear case. In particular, they involve usually a number of intrinsic parameters and interest centers not so much on determining a few specific solutions for fixed parameter-values but to assess the behavior of these solutions under general variations of the parameters. In structural analysis the parameters may characterize load points and load directions, material properties, geometrical data, etc., and the set of all solutions depending on these parameters has been called the equilibrium surface of the structure (see [37]). This equilibrium surface provides considerable insight into the behavior of the structure and its stability properties; we refer, for instance to [30], [38] for further discussions and various examples. From a numerical viewpoint the question then is to analyze computationally the shape and characterize features of this equilibrium surface. For this a principal tool is a general form of continuation process which allows a trace of a priori specified as well as intrinsically defined paths on the surface.

In the analysis of a structural problem by the finite element method we can compute only approximate points along a path on the solution surface of some discretized form of the original problem. Then -- in line with the earlier stated objectives of a practical computation -- we are faced with the need for assessing and controlling the errors along an entire segment of

such a path. In addition, since we are interested in specific features of the equilibrium surface, such as, the location of turning points, bifurcation points, stability boundaries, etc., we require also estimators of the error between such specific points and their computed approximations. Some a priori estimates of this type have been given recently in [16], [17], [18]. But, as in the linear case, we require, of course, computable and reliable a posteriori estimators which can then form the basis of adaptive procedures for computing with a prescribed accuracy the desired segments of solution paths on the equilibrium surface.

In the first part of this paper we consider a strongly nonlinear, one-dimensional rod-problem and show for it that effective a posteriori error estimates and adaptive procedures can be constructed and successfully integrated into a general continuation process for tracing paths on the equilibrium surface. In addition, we introduce error estimators for the computed location of turning points along such paths. Finally, various numerical results show the effectivity of all these approaches and ideas.

The results given here are as yet restricted in their applicability but should be capable of generalization to broader classes of problems. In order to illustrate where such generalizations should lead us we present in Part II of this paper an outlook on some further results for linear problems of the type we would hope to achieve ultimately also for the non-linear case. In particular, this includes the design and application of an adaptive finite element solver for certain two-dimensional elliptic problems and the extension of the use of error estimators and adaptive procedures to some parabolic problems. For further summaries of the specific results presented in this paper we refer to the outlines included at the beginning of each one of the two parts.

Part I: Nonlinear Problems

In this part, consisting of sections 2 to 9 of this paper, we discuss various aspects of the computational solution of a class of nonlinear problems in structural mechanics. More specifically, we consider as a model problem a planar, elastic rod involving both material as well as geometrical nonlinearities. But, as far as possible, the discussion is kept general and applicable to broader classes of problems. As indicated in the introduction, for nonlinear problems of this type we are interested in analyzing the shape and features of the equilibrium surface. Accordingly, we include here some general remarks about such surfaces and about the related computational problems involving solution paths, bifurcation and turning points, stability boundaries, etc.. Then a general continuation process is summarized and some of its extensions for determining limit points and tracing stability paths are discussed. In line with the comments in the introduction we introduce estimators for the error along the solution paths and of the computed critical values. Experimental results for the model problem show the effectivity of these estimators and of the adaptive procedures based upon them. Finally we discuss some aspects relating to the occurrence of so-called spurious solutions which occur frequently in discretized nonlinear problems.

2. A Model Problem

As a model problem we consider the planar deformations of a non-linearly elastic rod. More specifically, large deformations as well as material non-linearities are allowed. A comprehensive treatment of the global behavior of such rods was given in [1]. These rods can suffer not only flexure but also compression (or extension) and shear.

The basic configuration of the rod is sketched in Figure 1. We restrict ourselves to the case of a symmetric shape where it suffices to consider only one half of the rod, say, the left half in our picture. Then the configuration of the rod is described by two functions

$$\underline{r}: [0,1] \to R^2, \quad q: [0,1] \to R^2, \quad ||\underline{q}(s)||_2 = 1 \tag{2.1}$$

where $\underline{r}$ defines the axis of the deformed rod and $\underline{q}$ the direction of the cross-section (expressing the effect of the shear stresses). The parameter s represents the arc-length of the rod-axis in the straight reference configuration. Let e^1, e^2 be the global basis vectors as shown in the figure and u the angle between $\underline{q}$ and e^2. Then, with

$$U(u) = \begin{pmatrix} \cos u & \sin u \\ -\sin u & \cos u \end{pmatrix} \tag{2.2}$$

we define the local coordinate system by

$$\underline{p}(s) = U(u(s))^T e^1, \quad \underline{q}(s) = U(u(s))^T e^2 \tag{2.3}$$

and the strains ξ, η by

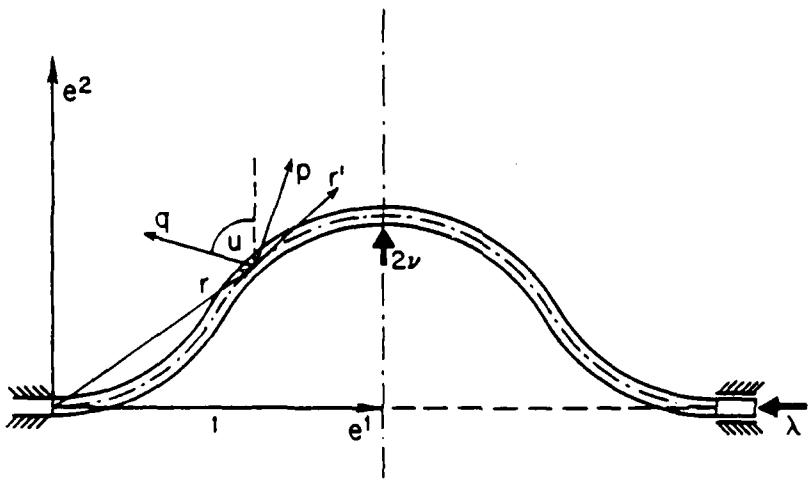


Figure 1

$$\underline{r}'(s) = U(u(s))^T \begin{pmatrix} 1 + \xi(s) \\ \eta(s) \end{pmatrix} . \tag{2.4}$$

Let $\underline{f}(s)$ and $\underline{m}(s)$ denote the resultant force and moment, respectively, in the cross-section at the point s. Since the load is constant, the equilibrium equations have the form

$$\underline{f}' = 0, \quad \underline{m}' + \underline{r}' \times \underline{f} = 0 \tag{2.5}$$

where primes denote differentiation with respect to s. Hence, we have

$$\underline{f}(s) = \binom{-\lambda}{\nu}, \quad \underline{m}(s) = M(s)(e^1 \times e^2) \tag{2.6}$$

and

$$0 = M'(s) + \binom{\nu}{\lambda}^T \underline{r}'(s) = M'(s) + \left[U(u(s)) \binom{\nu}{\lambda}\right]^T \begin{pmatrix} 1 + \xi(s) \\ \eta(s) \end{pmatrix} . \tag{2.7}$$

We introduce the constitutive equations

$$\begin{aligned} M(s) &= a(u'(s)), \\ \xi(s) &= b\left((e^1)^T U(u(s)) \binom{-\lambda}{\nu}\right) = b(-\lambda \cos u(s) + \nu \sin u(s)), \\ \eta(s) &= c\left((e^1)^T U(u(s)) \binom{\nu}{\eta}\right) = c(\nu \cos u(s) + \lambda \sin u(s)), \end{aligned} \tag{2.8}$$

where a, b, and c are given real functions. With the abbreviation

$$\hat{b}(u,\lambda,\nu) = (\lambda \sin u + \nu \cos u)(1 + b(-\lambda \cos u + \nu \sin u)) + (\lambda \cos u - \nu \sin u)\, c(\lambda \sin u + \nu \cos u) \tag{2.9}$$

our equation (2.7) now assumes the form

$$\frac{d}{ds} a(u') + \hat{b}(u,\lambda,\nu) = 0. \tag{2.10}$$

The fixed left endpoint and the symmetry requirement result in the homogeneous boundary conditions

$$u(0) = u(1) = 0. \tag{2.11}$$

Note that for $a(t) \equiv a_0 t$ and $b(t) \equiv c(t) \equiv 0$ the problem reduces to the classical Bernoulli-Euler problem

$$a_0 u'' + \lambda \sin u + \nu \cos u = 0, \quad u(0) = u(1) = 0. \tag{2.12}$$

Equation (2.10) is strongly nonlinear when $a(t) \not\equiv a_0 t$ and only mildly nonlinear when $a(t) = a_0 t$ as in (2.12). Mildly nonlinear problems of this form have been studied by many authors. For some recent results about finite element approximations for such equations see, for instance, [16], [17], [18] and the references given there.

In order to define the specific constitutive equations used in our numerical experiments let

$$g_1(t) = \arctan \tfrac{1}{2} t, \quad g_2(t) = \begin{cases} \dfrac{t}{1-t} & \text{for } t < 0 \\ t + 2t^2 & \text{for } t \geq 0 \end{cases}. \tag{2.13}$$

Then the three cases of (2.8) considered in this paper are as follows:

$$
\begin{array}{llllll}
 & \text{(i)} & \text{(Strongly nonlinear)} & a(t) = g_1(t), & b(t) = g_2(t), & c(t) = g_1(t) \\
(2.14) & \text{(ii)} & \text{(Mildly nonlinear)} & a(t) = \frac{1}{2}\,t, & b(t) = g_2(t), & c(t) = g_1(t) \\
 & \text{(iii)} & \text{(Classical Euler Rod)} & a(t) = \frac{1}{2}\,t, & b(t) \equiv 0, & c(t) \equiv 0
\end{array}
$$

Note that from (2.4) and (2.8) it follows that *the shape of the axis* of the rod is specified by

$$
\underline{r}(s) = \int_0^s U(u(\sigma))^T \begin{pmatrix} 1 + b(-\lambda \cos u(\sigma) + \nu \sin u(\sigma)) \\ c(\lambda \sin u(\sigma) + \nu \cos u(\sigma)) \end{pmatrix} d\sigma .
$$

Moreover, by (2.7) the moment balance is given by

$$
(2.16) \qquad M(0) - M(1) = \binom{\nu}{\lambda}^T \underline{r}(1).
$$

We refer to [1] for further theoretical results about this problem.

3. Remarks on Equilibrium Surfaces

Our model problem (2.10)/(2.11) involves the two parameters λ, ν representing the components of the resultant force. For any fixed λ, ν a solution $u = u(s)$ of the boundary value problem defines an equilibrium configuration of the rod. For practical applications it is rarely sufficient to determine only the equilibria for a few given loads. Instead we are interested usually in the changes of the equilibrium position under specific changes of the load.

Frequently in structural mechanics for a fixed direction the load-intensity is used as a parameter. For example, in our case we may consider the family of loads $-\lambda e^1$ acting in the direction of the axis of the rod in its undeformed configuration. If λ varies say from $\lambda = 0$ to some maximal value we obtain in this way a curve of equilibria of the rod. Often this curve is parametrized in terms of λ. But at turning points this parametrization breaks down and hence it is usually better to consider the curve in the form $\lambda = \lambda(t)$, $u = u(t)(s)$ with another, more suitable parameter t. Clearly, for different load directions or starting points we obtain a different curve.

In our case the problem depends on the two parameters λ and ν. Accordingly, it turns out that all these curves form a two-dimensional surface in (λ,ν,u)-space, the so-called equilibrium surface of the problem. In general, this surface has a complicated shape but its consideration provides considerable insight into the equilibrium behavior of our structure.

In order to illustrate these concepts it may be helpful to consider for a moment a particular frame-work model of our rod (see also [30], p. 291). Two rigid rods of length 1 each are flexibly jointed as shown in Figure 2 and a linear spring with spring-constant k is trying to hold the structure

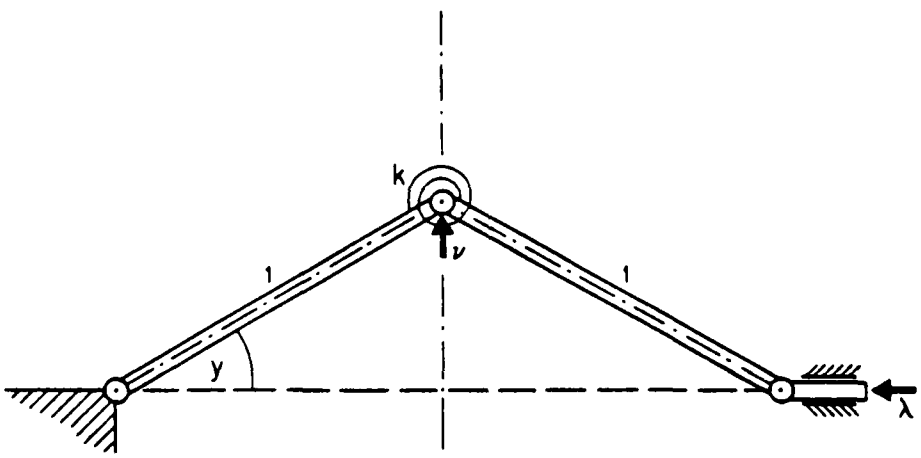


Figure 2

in its straight reference configuration. For simplicity, let $k = 1/2$. Any equilibrium position is characterized by the value of the angle y, and it is easily seen that the equilibrium equation has the form

$$y - \lambda \sin y - \frac{1}{2} \nu \cos y = 0. \tag{3.1}$$

Hence, the equilibrium surface here is a two-dimensional surface in the three-dimensional space with coordinates λ, ν, y. It is a ruled surface for which the generators have constant values of y. Figure 3 shows level lines of a projection of part of this surface in the direction of the y-axis. In addition, Figures 4 and 5 illustrate the cross sections of the surface with planes $\lambda = \text{const}$ and $\nu = \text{const}$, respectively. The latter two figures show clearly that the point $\lambda = 1$, $\nu = 0$, $y = 0$ is special. More precisely, we have here a bifurcation point where two equilibrium curves for the load direction $(1,0)^T$ are intersecting. The particular shape of the surface near this point represents a so-called cusp-catastrophe. The reason for this name is obvious from Figure 3.

Evidently, the equilibrium surface characterizes much better the complete behavior of the framework than the usual traces of the equilibrium-curves corresponding to fixed load directions. Furthermore, the surface may be partitioned into subsets of stable and instable equilibria. As usual, an equilibrium-solution is called (locally) stable if it represents a local minimum of the total energy. In our case, it is readily seen that instability occurs when

$$1 - \lambda \cos y + \frac{1}{2} \nu \sin y < 0. \tag{3.2}$$

In Figure 3 the upper part of the visible surface consists of stable equilibria.

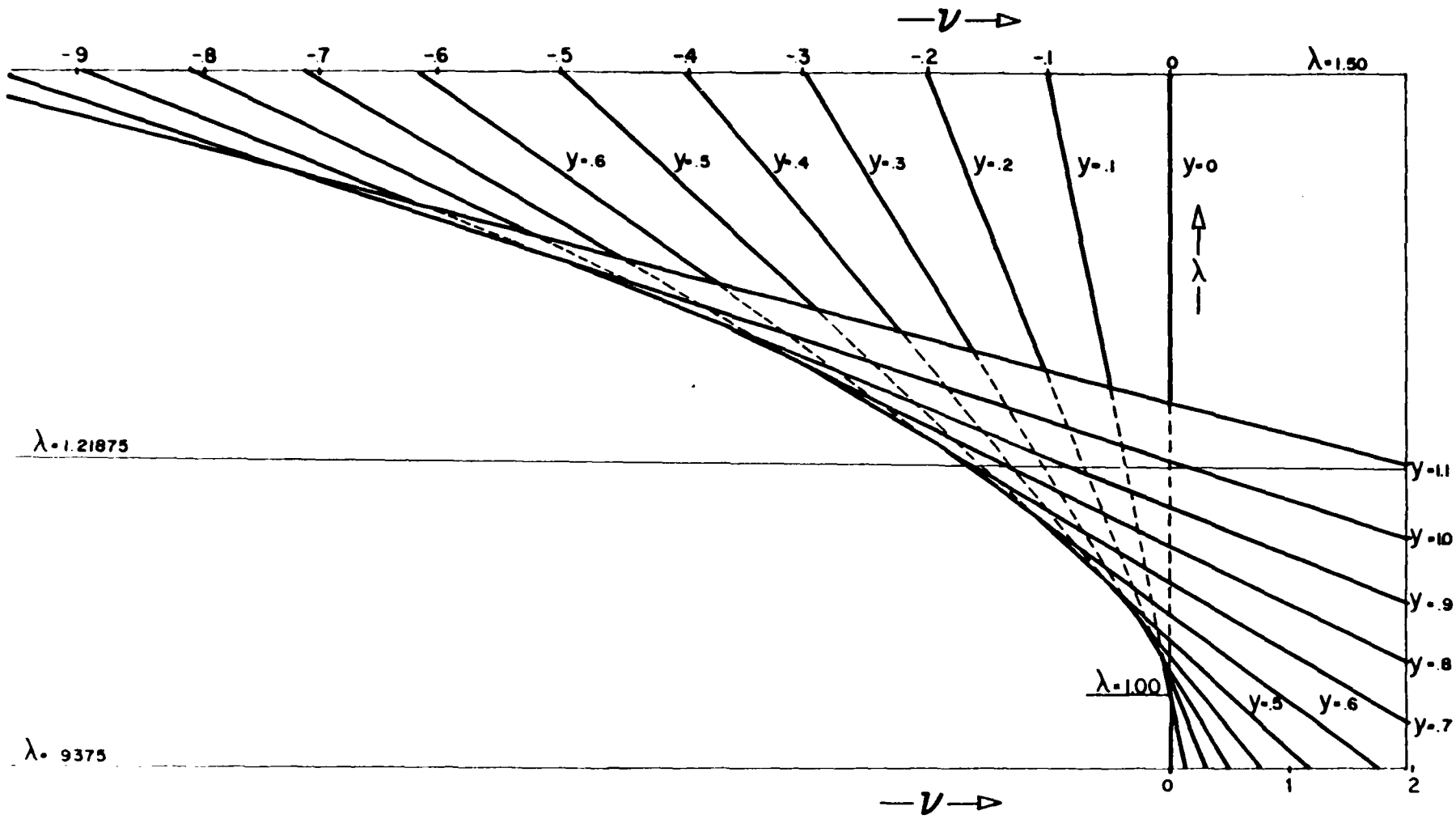

—ν→
-9
-8
-7
-6
-5
-4
-3
-.2
-.1
0
λ=1.50
y=.6
y=.5
y=.4
y=.3
y=.2
y=.1
y=0
λ
λ=1.21875
y=1.1
y=1.0
y=.9
y=.8
y=.7
λ=1.00
y=.5
y=.6
λ= .9375
0
1
2
—ν→

Figure 3

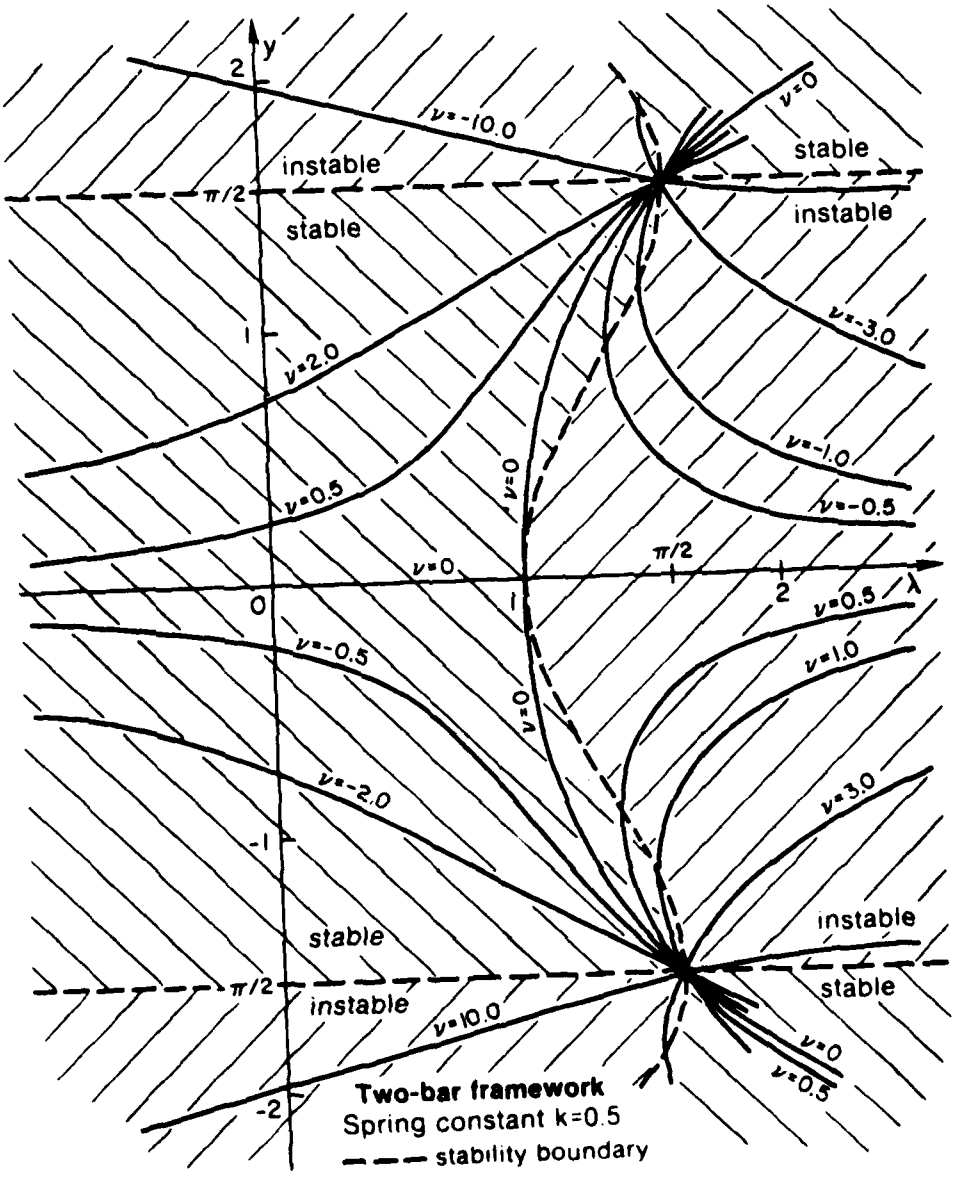

y
2
1
0
-1
-2
π/2
π/2
1
2
λ
ν=-10.0
ν=0
stable
instable
instable
stable
ν=-3.0
ν=2.0
ν=-1.0
ν=0.5
ν=-0.5
ν=0
ν=0
ν=0.5
ν=-0.5
ν=1.0
ν=0
ν=-2.0
ν=3.0
instable
stable
stable
instable
ν=10.0
ν=0
ν=0.5
Two-bar framework
Spring constant k=0.5
stability boundary

Figure 4

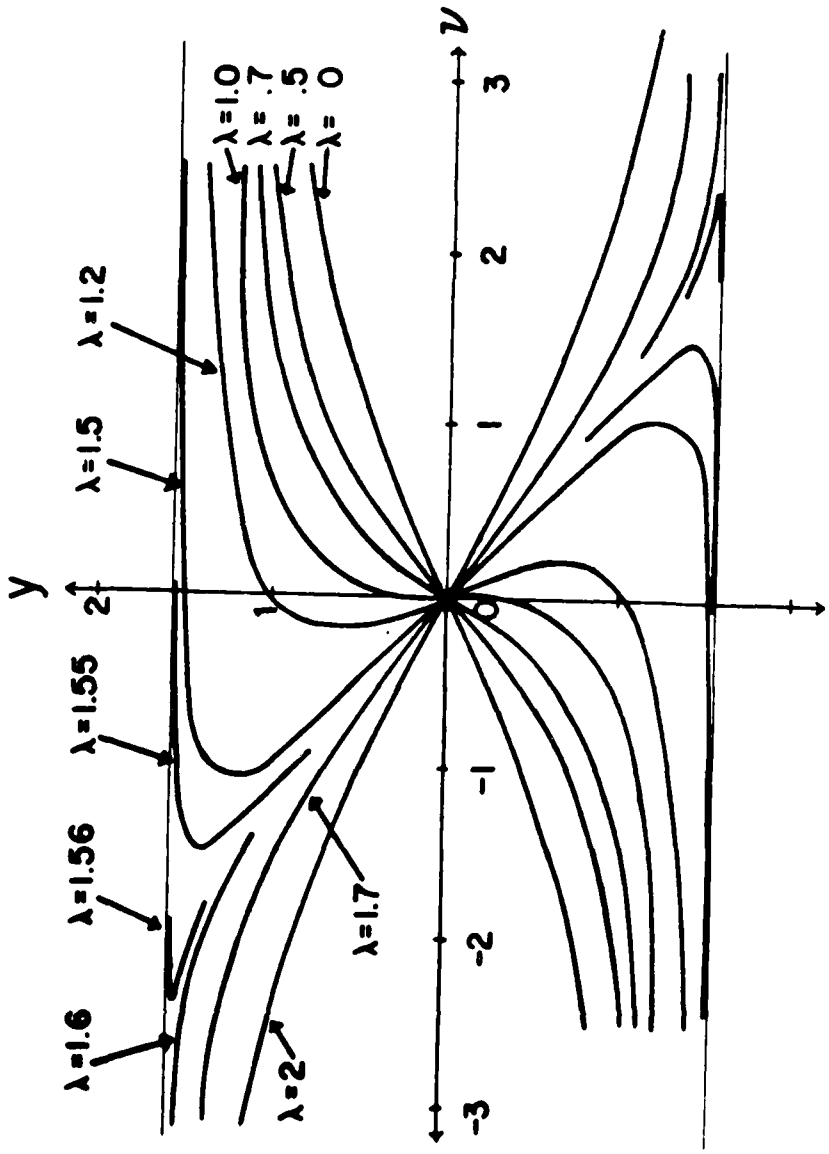


Figure 5

The boundary of the stability region is the envelope of the generators. This stability boundary is shown as a dashed line in Figure 4. It consists of the turning points with respect to the load parameter λ on the curves for constant values of ν. These points represent the buckling points of the structure for the particular values of ν. Thus, at one glance the surface provides us with information about the stable equilibria, buckling loads, etc. for our framework.

This indicates that for a deeper understanding of the behavior of a structure we should analyze the form of the equilibrium surface. Of course, the choice of the parameters λ and ν entering into the definition of our surface is basically arbitrary albeit natural. Practical considerations show that the load-component λ is essential for the study of the buckling behavior of the framework. At the same time it is necessary to introduce one or several parameters describing possible perturbations. There are theoretical reasons to consider a particular minimal number of such perturbation parameters, but we shall not enter into that question here.

In view of these observations the aim of our analysis is to determine the form of the equilibrium surface. In our simple example this is easy since no discretization is involved. But in general, we can compute only points which are approximately on the solution surface of a discretization of the original problem. In line with the earlier stated requirements for any practical computation this raises the question of assessing the errors committed here.

Analogous to the framework the equilibrium surface of the rod-problem (2.10) is a two-dimensional surface in the three-dimensional space with coordinates λ, ν, u. Suppose that $u(\lambda,\nu)$ and $\tilde{u}(\lambda,\nu)$ denote the exact solution and the computed, approximate solution, respectively, for

given parameter values λ, ν. An obvious possibility for the error assessment then is the requirement that $\|u(\lambda,\nu) - \tilde{u}(\lambda,\nu)\| \leq \varepsilon$. However, this is reasonable only if both u and $\tilde{u}$ exist for the particular parameter values. This is by no means guaranteed as Figure 6 shows which provides a sketch of possible curves $u(\cdot,\nu_0)$ and $u(\cdot,\nu_0)$. For $\lambda < \lambda_T$ both u and $\tilde{u}$ exist but for $\lambda > \lambda_T$ there is no value of $u(\lambda,\nu_0)$ which can be compared with the computed value $\tilde{u}(\lambda,\nu_0)$. Hence, it will be necessary to consider different measures of the error between the exact and computed solutions. A possibility is to ask for an estimate which guarantees that for any computed $\tilde{u}(\lambda,\nu)$ there exists an exact solution $u(\tilde{\lambda},\tilde{\nu})$ with

$$|\lambda - \tilde{\lambda}|^{\alpha} + |\nu - \tilde{\nu}|^{\beta} + \|\tilde{u}(\lambda,\nu) - u(\tilde{\lambda},\tilde{\nu})\| \leq \varepsilon \tag{3.3}$$

and suitable exponents α, β. Clearly, this avoids the difficulty with the earlier type of estimate. In addition to estimates such as (3.3) other more specific error-assessments are often desired. For example, it is frequently of considerable practical interest to measure the error $|\lambda_T - \tilde{\lambda}_T|$ in the parameter values corresponding to the turning points of the exact and computed solutions.

In connection with this discussion it should be noted also that the equilibrium surface of the discretized problem may consist of more connected components than the surface of the exact problem. In other words, there may be numerical solutions which do not approximate exact solutions. Such spurious solutions have been observed in many contexts (see eg. [20], [28] and the references given there). From our viewpoing the question then is to identify the segments of solution paths on the equilibrium surface of the discrete problem for which the error estimates are below an acceptable threshold.

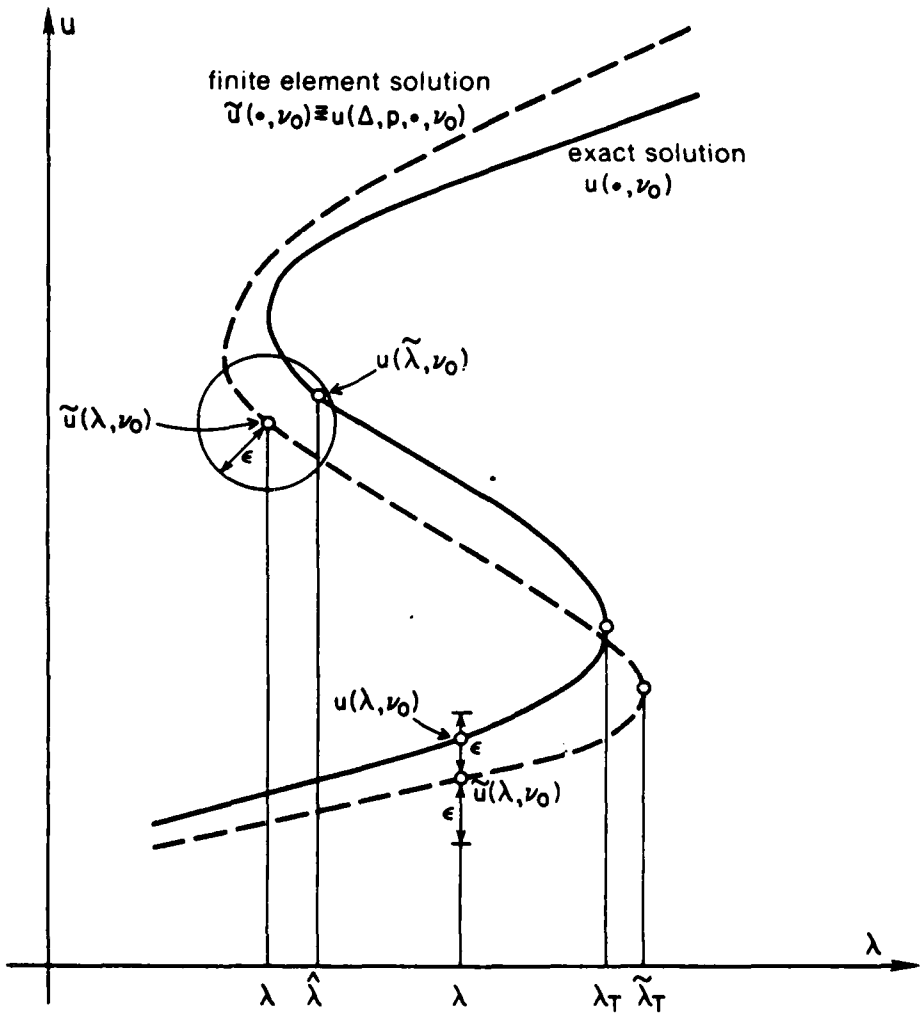

u
finite element solution
$\tilde{u}(\cdot,\nu_0) \equiv u(\Delta,p,\cdot,\nu_0)$
exact solution
$u(\cdot,\nu_0)$
$u(\tilde{\lambda},\nu_0)$
$\tilde{u}(\lambda,\nu_0)$
ϵ
$u(\lambda,\nu_0)$
ϵ
$\tilde{u}(\lambda,\nu_0)$
ϵ
λ
λ
$\hat{\lambda}$
λ
λ_T
$\tilde{\lambda}_T$

Figure 6

4. A Finite Element Approximation of the Rod Problem

We return again to the rod problem (2.10)/(2.11). The finite element solution of this boundary value problem is based on the standard variational form

$$\int_0^1 [-a(u')v' + \hat{b}(u,\lambda,\nu)v]ds = 0 \tag{4.1}$$

where v varies over a given space of test functions.

A mixed finite element method was used to approximate the solutions of (4.1). More specifically, let

$$\Delta: \quad 0 = s_0 < s_1 < s_2 < \ldots < s_n = 1, \quad n = n(\Delta) \tag{4.2}$$

be a given partition of the interval $[0,1]$. Then on the i-th interval $I_i = [s_{i-1},s_i]$, $1 \le i \le n$, of Δ the solution is approximated by

$$u_i(s) = \sum_{j=0}^{p} y_{ij}\, \psi_j(\tau_i(s)), \quad \forall\, s \in I_i, \quad 1 \le i \le n, \tag{4.3}$$

where $\psi_0,\psi_1,\ldots,\psi_p$ are the standard Legendre polynomials on $[-1,+1]$ and

$$\tau_i(s) = \frac{2}{h_i}\left(s - \frac{1}{2}(s_{i-1} + s_i)\right), \quad h_i = s_i - s_{i-1}, \quad 1 \le i \le n. \tag{4.4}$$

At the interior nodes of Δ the functions and their derivatives up to order $d-1$ are to match continuously $(2d \le p+1)$. Thus, together with the two

boundary conditions at s_0 and s_n we have to add $d(n-1) + 2$ constraints to (3.1). In order to take account of the corresponding constraints for the test functions an equal number of Lagrange multipliers have to be introduced. Hence, altogether, we have $(p+1)n + d(n-1) + 2$ variables. In the standard finite element method the use of the corresponding elements results in

$$(p+1)n - d(n-1) - 2, \tag{4.5}$$

degrees of freedom. This number is used later in all estimations of the total work.

The mixed method of this form was selected for our example computations because it allows for a flexible use of elements of various degree and smoothness. With this it was possible to assess among other points the amount of work for a fixed accuracy corresponding to different degrees of approximation and varying smoothness. High order elements are used frequently, for instance, in the P-version of the finite element method, (see eq. [14], [15], [41]).

5. A Continuation Method

In structural mechanics continuation methods are usually viewed in a narrow sense as methods for following numerically a specific load curve parametrized by a load intensity. As we discussed in section 3 the solution set of structural problems generally constitutes some surface or manifold. For example, for our rod-problem (2.10)/(2.11) the equilibrium surface is a two-dimensional surface in (λ,ν,u)-space. For a thorough understanding of the equilibrium behavior of a non-linear structure we are interested in determining the principal features of the equilibrium surface, as, for example, the location of critical boundaries and bifurcation points where stability is lost, etc..

For a numerical analysis of a given equilibrium surface we need to consider continuation methods in a broader sense as a collection of numerical procedures for accomplishing tasks of the following type:

(5.1)

(i) Follow numerically any curve on the surface specified by a particular combination of parameter values.

(ii) Determine on such a curve the critical points where stability may be lost.

(iii) From any such critical point follow a path in the critical boundary.

(iv) On any curve specified by (i) determine the location of bifurcation points and the paths intersecting at that point.

Methods for these and related tasks have been developed by various authors in recent years (see eg. [23], [24], [27], [29], [31], [32], [34], [39], [40] and the references given there). We shall not go into details here but sketch only some of the principal features of these methods as they apply in the present setting.

Generally, a (finite-element) discretization of a structural problem leads to a finite dimensional non-linear equation of the form

$$F(y,p) = 0 \tag{5.2}$$

where $y \in R^n$ is a vector of state variables (eg., deformations), $p \in R^m$ a vector of parameters, and $F: R^n \times R^m \to R^n$ a given function. For ease of discussion we shall assume here that $m = 2$. The set

$$E(F) = \{(y,p) \in R^n \times R^2 \mid F(y,p) = 0\} \tag{5.3}$$

of all solutions of (5.2) is the equilibrium surface of our discretized problem.

Let $DF(y,p)$ denote the derivative of F; that is, the $n \times (n+2)$-(Jacobian) matrix of all partial derivatives of F. We assume that the so-called regularity set

$$R(F) = \{(y,p) \in R^n \times R^2 \mid \text{rank } DF(y,p) = n\} \tag{5.4}$$

is not empty. For example, in the case of the framework of Figure 2 we have $y \in R^1$, $p = (\lambda,\nu)^T \in R^2$ and

$$DF(y,p) = (4k - 2\lambda \cos y + \nu \sin y, -2 \sin y, -\cos y). \tag{5.5}$$

The three partial derivatives in this 1×3 matrix $DF(y,p)$ can never be zero at the same time. Hence, the rank of that matrix is always one and

the regularity set $R(F)$ is the entire space R^3.

On the equilibrium surface (5.3) we may define a curve by specifying some relation between the two components λ, ν of the parameter vector $p = (\lambda,\nu)^T \in R^2$. For example, suppose that we fix a linear relation of the form

$$q^T(p-p^o) \equiv q_1(\lambda-\lambda^o) + q_2(\nu-\nu^o) = 0 \tag{5.6}$$

where $q = (q_1,q_2)^T$ is a given non-zero vector and $p^o = (\lambda^o,\nu^o)$ a specified vector of parameter values. Hence, we are now interested in the solutions of (5.2) for which (5.6) holds; in other words, we are considering the expanded equation

$$\hat{F}(y,p) \equiv \begin{pmatrix} F(y,p) \\ q^T(p-p^o) \end{pmatrix} = 0. \tag{5.7}$$

In our examples the parameter vector p represents a load. Hence, p^o is a reference load, (5.6) defines a load direction and the solutions of (5.7) are the equilibria of the structure in this particular load direction.

The derivative of $\hat{F}$ is the $(n+1) \times (n+2)$ matrix

$$D\hat{F}(y,p) = \begin{pmatrix} D_yF(y,p) & D_p(y,p) \\ 0 & q^T \end{pmatrix} \tag{5.8}$$

and it is readily seen that the regularity set $R(\hat{F})$ of $\hat{F}$ is a subset of $R(F)$; that is

$$(5.9) \qquad R(\hat{F}) \subset R(F).$$

It is important to note that in general there are points in $R(F)$ which do not belong to $R(\hat{F})$. For example, for the framework of Figure 2 suppose that we are interested in tracing a curve defined by $\nu = \nu^0$. Hence, in (5.6) we use $q = (0,1)^T$, $p^0 = (0,\nu^0)^T$, and (5.8) has the form

$$(5.10) \qquad D\hat{F}(y,p) = \begin{pmatrix} 2 - 2\lambda \cos y + \nu \sin y & -2 \sin y & -\cos y \\ 0 & 0 & 1 \end{pmatrix}.$$

The matrix has rank two everywhere in R^3 except on the straight lines defined by $\lambda = (-1)^i$, $y = i\pi$, $i = 0,\pm 1,\pm 2,\ldots$ and variable ν. Thus, $R(\hat{F})$ is no longer equal to R^3 as was the case with $R(F)$.

The intersection of the equilibrium surface $E(F)$ with the straight lines where $D\hat{F}(y,p)$ has only rank one are the points with the coordinates $\lambda = (-1)^i$, $\nu = 2(-1)^{i+1} i\pi$, $y = i\pi$, $i = 0,\pm 1,\pm 2,\ldots$. In the region where $|y| \leq \pi/2$ the only such point is the bifurcation point $\lambda = 1$, $\nu = 0$, $y = 0$ of Figure 4. Note that this point became a singular point only by considering the family of curves $\nu = \nu^0$ on the equilibrium surface. Different families of curves generate different singular points. For instance, if we trace the family of curves $\lambda = \lambda^0$ for given λ^0, then in our strip $|y| \leq \pi/2$ the only points not in $R(\hat{F})$ have the coordinates $\lambda = \pi/2$, $y = \pm \pi/2$, $\nu = -2$. This effect is clearly visible in Figures 4 and 5.

For a fixed vector q the points of the regularity set $R(\hat{F})$ are of principal interest for our continuation processes, since, under simple conditions about F, there exists a unique solution curve of (5.7) through

any such point. More precisely we have the following existence result (see [32]):

<u>Theorem</u>: Let $F: R^n \times R^2 \to R^n$ be continuously differentiable and DF locally Lipschitzian. With given $q, p^0 \in R^2$ form the expanded mapping $\hat{F}$ of (5.7). Then for any $(y^0, p^0) \in R(\hat{F})$ there exists a unique, continuously differentiable solution $y: J \subset R^1 \to R^n$, $p: J \subset R^1 \to R^2$ such that

$$\hat{F}(y(s),p(s)) = 0, \quad \forall s \in J, \quad y(0) = y^0, \quad p(0) = p^0$$

(5.11)

$$(y(s),p(s)) \in R(\hat{F}), \quad ||(y'(s),p'(s))||_2 = 1, \quad \forall s \in J$$

where the open interval $J \subset R^1$ is maximal under set-inclusion. Moreover, if $\tilde{s}$ is a finite endpoint of J then for $s \to \tilde{s}$, $s \in J$ either the points $(y(s),u(s))$ tend to the boundary of $R(\hat{F})$ or $||y(s),u(s))||_2$ goes to infinity.

The effect of this theorem is clearly visible in Figure 4. Through any point in the shown region, except the bifurcation point $\lambda = 1$, $\nu = 0$, $y = 0$, there is exactly one solution curve which terminates only at a bifurcation point. Note that at $\lambda = \pi/2$, $y = \pm \pi/2$ the $\nu = \nu^0$ curves in R^3 do not intersect although in the figure this appears to happen.

For the discussion of a continuation process for computing a curve (5.11), it is useful to use the abbreviation $x = (y,u)$ for the vectors $(y,u) \in R^{n+2}$. It can be shown that on the regularity set $R(\hat{F})$ there exists a unique tangent mapping T such that

$$(5.12) \qquad T: R(\hat{F}) \to R^{n+2}, \quad D\hat{F}(x)Tx = 0, \quad ||Tx||_2 = 1, \quad \det\begin{pmatrix} D\hat{F}(x) \\ (Tx)^T \end{pmatrix} > 0.$$

In other words, for any point $x \in R(\hat{F})$ the vector Tx of length one is tangent to the unique solution curve through that point. The determinant condition in (5.12) was introduced to fix the orientation of that tangent direction.

Most modern numerical continuation processes for tracing a curve (5.11) are of the predictor-corrector type. But the methods differ in the choice of the parametrization of the curve. We shall discuss here a locally parametrized process of the form presented in [32], [34].

Let $x^0, x^1, \ldots, x^k$, $k \geq 0$, be the approximate points along the curve which have been computed so far. The computation of the next point x^{k+1} has schematically the following form:

(5.13)

1. At x^k compute the tangent vector Tx^k.

2. Choose the index $i = i_k$, $1 \leq i \leq n+2$, of the variable x_i which is to serve as the current continuation parameter.

3. Choose a steplength t_{k+1} along the Euler line $x^k + tTx^k$.

4. Compute the predicted point $x^{pred} = x^k + t_{k+1}\, Tx^k$ and the corresponding "corrector-constant" γ_k.

5. With x^{pred} as starting point apply Newton's method to the corrector equation

$$Gx = \begin{pmatrix} \hat{F}x \\ (e^{i_k})^T (x - x^{pred}) - \gamma_k \end{pmatrix} = 0.$$

If "no convergence" then replace t_{k+1} by $\frac{1}{2} t_{k+1}$ and go to 4.
If "convergence" then use the last iterate as the next continuation point x^{k+1}.

For the computation of the tangent vector Tx^k note that

$$(5.14) \qquad \det \begin{pmatrix} D\hat{F}(x^k) \\ (e^i)^T \end{pmatrix} = [(e^i)^T Tx^k] \det \begin{pmatrix} D\hat{F}(x^k) \\ (Tx^k)^T \end{pmatrix}, \quad i = 1,\ldots,n+2$$

where $e^1,\ldots,e^{n+2}$ are the natural basis vectors of R^{n+2}. Hence, with any index i such that $(e^i)^T Tx^k \neq 0$ the tangent vector may be computed as follows

$$(5.15) \qquad \begin{pmatrix} D\hat{F}(x^k) \\ (e^i)^T \end{pmatrix} v = e^{n+2}$$

$$Tx^k = \sigma \frac{v}{||v||_2}, \quad \sigma = [\text{sign} \det \begin{pmatrix} D\hat{F}(x) \\ (e^i)^T \end{pmatrix}]/[\text{sign } (e^i)^T Tx^k].$$

For $k \geq 1$ we use the index $i = i_{k-1}$ of the continuation variable during the previous step while for $k = 0$ the index of the original, user-provided continuation variable is taken.

In line with (5.14) the index $i = i_k$ of the variable x_i which is to serve as the current continuation variable is chosen such that

$$(5.16) \qquad |(e^{i_k})^T Tx^k| = \max_{i=1,\ldots,n+2} |(e^i)^T Tx^k|.$$

This represents the variable for which the basis vector e^i forms the smallest angle with the tangent Tx^k. Note that -- unless the step from x^{k-1} to x^k was unreasonably large -- the choice of $i = i_{k-1}$ in (5.15) should guarantee that the matrix indeed is nonsingular.

In order to determine a steplength we proceed as in the case of ODE-

solvers and introduce the quadratic polynomial

$$q(t) = x^k + t\,Tx^k + t^2\,w^k \tag{5.16}$$

with

$$w^k = \frac{1}{h_k}\left(Tx^k - \frac{1}{h_k}(x^k - x^{k-1})\right), \quad h_k = ||x^k - x^{k-1}||_2 \tag{5.17}$$

for which

$$q(0) = x^k, \quad q'(0) = Tx^k, \quad q(-h_k) = x^{k-1}. \tag{5.18}$$

The distance between x^{pred} and the curve is to be smaller than some tolerance $\varepsilon_{k+1} > 0$. Asymptotically for sufficiently small h_k and t_{k+1} this requirement may be approximated by

$$||x^{pred} - q(t_{k+1})||_2 \leq \varepsilon_{k+1} \tag{5.19}$$

whence

$$t_{k+1} \leq \sqrt{\frac{\varepsilon_{k+1}}{||w^k||_2}}\ . \tag{5.20}$$

The polynomial (5.16) is numerically ill-defined when the angle

$$\alpha_k = \arccos\left[(Tx^k)^T\left(\frac{1}{h_k}(x^k - x^{k-1})\right)\right] \in [0,\pi] \tag{5.21}$$

is small, that is, when the path is fairly straight. This makes it advisable not to attempt to use $q(t_{k+1})$ in place of x^{pred} as the predicted point. The norm of w^k may be evaluated easily in the form

$$(5.22) \qquad ||w^k||_2 = \frac{2}{h_k} \left|\sin \frac{\alpha_k}{2}\right| .$$

The choice of the "corrector-constant" γ_k in line 4 of (5.13) defines the point on the path (5.11) which is to be approximated by the corrector iteration. More specifically, suppose that x^k is an approximation of the point $x(s_k)$ on the curve, then for given γ_k and sufficiently small t and Δs the equation

$$(5.23) \qquad (e^{i_{k+1}})^T (x(s_k+\Delta s) - (x^k+t\,Tx^k)) = \gamma_k$$

defines a one-to-one correspondence between t and Δs. Evidently, the choice $t = \Delta s$ underlying the inequality (5.19) is a natural but by no means required goal. If we replace $x(s)$ in (5.23) by $q(t)$ then $\Delta s = t_{k+1}$ implies the use of the corrector constant

$$(5.24) \qquad \gamma_k = \frac{t_{k+1}^2}{h_k} [(e^{i_k})^T Tx^k - \frac{1}{h_k} (e^{i_k})^T (x^k-x^{k-1})].$$

Obviously, γ_k should be evaluated in double-precision to reduce the possible effect of subtractive cancellation when the curve is nearly straight.

For the definition of the steplength this leaves the choice of the tolerance ε_{k+1} in (5.20). A study of some aspects of this question has been given in [19]. For our discussion here it suffices to mention only a

simple but rather effective technique.

For any real numbers $a < b$ we define the piecewise linear function

$$\mu(t;a,b) = \begin{cases} a & \text{if} \quad t \le a \\ t & \text{if} \quad a \le t \le b \\ b & \text{if} \quad t \ge b \end{cases} . \tag{5.25}$$

The steplength calculation uses as input the angle α_k of (5.21), the prediction error

$$\delta_k = ||(x^{k-1}+t_k\, Tx^{k-1}) - x^k||_2 \tag{5.26}$$

at the last step, and the last step h_k defined in (5.17). With these quantities the calculation is essentially based on (5.20) and proceeds as follows:

(5.27)

1. If $\alpha_k \le \alpha_{min}$ then $\theta := \kappa$, go to 4 [straight line case]
   else if $\alpha_k \ge \alpha_{max}$ then $\theta := \frac{1}{\kappa}$, go to 4 [hairpin case]
2. $\bar{\delta} := \mu(\frac{\delta_k}{h_k}; \delta_{min}, \delta_{max})$ [adjustment of relative tolerance]
3. $\theta^2 := \mu(\frac{\bar{\delta}}{2|\sin \frac{1}{2}\alpha_k|}, \frac{1}{\kappa^2}, \kappa^2)$ [adjustment of θ]
4. $t_{k+1} := \mu(\theta h_k, t_{min}, t_{max})$ [adjustment of step]

Here $0 < \alpha_{min} < \alpha_{max} < \pi$ are bounds for the angle α_k and κ is the maximally allowable step-increase, that is, we want to ensure that the step-factor θ is bounded by

$$\frac{1}{\kappa} \leq \theta \leq \kappa. \tag{5.28}$$

In line with this the tolerance bounds in line 2. of (5.27) are defined by

$$\delta_{min} = \frac{2}{\kappa^2} \sin \frac{1}{2} \alpha_{max}, \quad \delta_{max} = 2\kappa^2 \sin \frac{1}{2} \alpha_{min}. \tag{5.29}$$

The condition $\delta_{min} < \delta_{max}$ places some restriction on the choice of α_{min}, α_{max} and κ. Experience has shown that $\alpha_{min} = 0.05$, $\alpha_{max} = \pi/2$, and $\kappa = 3$ are effective choices for a large range of problems. The actual step is enforced to fall between the bounds $0 < t_{min} < t_{max}$. Here t_{min} depends on the accuracy of the machine and t_{max} reflects the desired density of points along the curve.

Line 5. of (5.13) requires some provisions for monitoring the convergence of the corrector iteration. For Newton-type methods it has been found satisfactory to declare non-convergence if either one of the three conditions is true

$$\begin{aligned} &\text{(i)} \quad ||Gx^{k,j}|| \geq \theta ||Gx^{k,j-1}||, \quad \text{for some } j \geq 1 \\ &\text{(ii)} \quad ||x^{k,j}-x^{k,j-1}|| \geq \rho ||x^{k,j-1}-x^{k,j-2}||, \quad \text{for some } j \geq 2 \\ &\text{(iii)} \quad j \geq j_{max} \end{aligned} \tag{5.30}$$

where $\{x^{k,j}\}$ denotes the sequence of corrector iterates, ϕ is a suitable constant, say $\phi = 1.05$, and j_{max} an integer, say, $j_{max} = 5$. On the other hand, the iterate $x^{k,j}$ is accepted as the next point x^{k+1} if

$$(5.31) \qquad (||Gx^{k,j}|| \leq \varepsilon_1) \text{ and } (||x^{k,j}-x^{k,j-1}|| \leq \varepsilon_2 + \varepsilon_3||x^{k,j}||), \text{ for some } j \geq 1$$

with given tolerances $\varepsilon_1, \varepsilon_2, \varepsilon_3$ which depend on the machine and the problem.

As stated in (5.13), in the case of non-convergence the predictor step is halved and the corrector is restarted with the corresponding value (5.24) of the corrector-constant γ_k. Of course, here it is required that $\frac{1}{2} t_{k+1} \geq t_{min}$ else some user dependent action is required. On the other hand, if convergence is declared with $x^{k+1} = x^{k,j}$ as the last iterate then a control program has to decide whether a new predictor/corrector step is needed or not.

The algorithm described here has been used extensively and with excellent success on a wide range of problems. In particular, it is well suited for the task (i) of (5.1). For practical use it is desirable to incorporate special features which allow, for instance, for the detection and computation of any target point on the curve where a specific variable has a given value, or for the detection and computation of a limit point. The latter feature will be discussed briefly in the next section.

6. Determination of Limit Points and Critical Boundaries

In this section we discuss briefly some of the algorithmic aspects relating to the tasks (ii) and (iii) of (5.1), that is, to the determination of critical points on a continuation curve and the trace of a path in the critical boundary. As before let (5.2) be the given system of non-linear equations which defines the equilibrium surface $E(F)$ of (5.3) for our discretized problem.

Suppose that there exists a total energy-functional $\pi: R^n \times R^2 \to R^1$ with $D_y\pi(y,p)^T = -F(y,p)$, $\forall\, y \in R^n$, $p \in R^2$. Then an equilibrium $(y^o,p^o) \in E(F)$ is stable if $\pi(y,p^o)$ has a proper local minimum at $y = y^o$ and hence if $D_yF(y^o,p^o)$ is negative definite. Let

$$C(F) = \{(y,p) \in R^n \times R^p \mid D_yF(y,p) \text{ is singular}\}; \tag{6.1}$$

then the so-called critical boundary $C(F) \cap E(F)$ contains the (relative) boundary of the set of stable equilibria on the equilibrium surface $E(F)$, although the two sets are not necessarily equal. The tasks (5.1) (ii) and (iii) concern the computation of points and curves in the critical boundary $C(F) \cap E(F)$. Note that this set is also well-defined if there is no total energy functional π. But then $D_yF(y,p)$ is non-symmetric and stability may be lost at points not in $C(F) \cap E(F)$.

As in the previous section we consider curves in $E(F) \cap R(F)$ defined by an augmented equation (5.7). As we saw before the regularity set $R(\hat{F})$ of the expanded function $\hat{F}$ is contained in $R(\hat{F})$. More specifically, it can be shown that

$$R(\hat{F}) = [R(F) \setminus C(F)] \cup L(F,q) \tag{6.2}$$

where

$$(6.3) \qquad L(F,q) = \{(y,p) \in R^n \times R^2 \mid \dim \ker D_yF(y,p) = 1,\ D_pF(y,p) \notin \operatorname{rge} D_yF(y,p)\}.$$

In other words, either a point $(y,p) \in R(\hat{F})$ on our curve does not belong to $C(F)$ or it is in the set $L(F,q)$. As the notation already indicates this set $L(F,q)$ depends only on F and the vector q used in (5.6) to specify the curve. Moreover, one can show that

$$(6.4) \qquad L(F,q) = \{x \in R(\hat{F}) \mid (e^j)^T Tx = 0, \quad j = n+1, n+2\}.$$

In other words, the points $(y,p) \in L(F,q)$ on our curve are exactly those points of $R(\hat{F})$ where the tangent of the curve is orthogonal to the 2-dimensional subspace of the parameter variables. Thus, in standard terminology, the points of $L(F,q)$ are the limit points of $\hat{F}$ in $R(\hat{F})$ or more precisely the limit points of F with respect to the direction q.

Suppose now that our continuation process of Section 5 is applied to trace the curve defined by (5.7) and a specific starting point $x^0 = (y^0,p^0) \in R(\hat{F})$ for which $\hat{F}x^0 = 0$. As shown in [32] the computation will continue as long as the curve remains in $R(\hat{F})$. In order to detect on the curve a limit point of $\hat{F}$ we have to look for points of $L(F,q)$. For this it is only necessary to monitor one component $(e^\ell)^T Tx$ of the tangent vector Tx along the curve, provided the index ℓ, $n+1 \le \ell \le n+2$, was chosen such that the vector q has a non-zero ℓ-th component; that is,

$$(6.5) \qquad q^T e^\ell \neq 0.$$

This follows readily from the characterization of $L(F,q)$ and the form of $\hat{F}$.

If our starting point $x^0 \in R(\hat{F})$ represented a stable equilibrium then the first limit point encountered on our path indicates a point where stability may be expected to be lost. For structural problems this is typically a point where buckling may occur. Let $x = x(s) \in R(\hat{F})$, $s \geq 0$, $x(0) = x^0$, be the segment of the exact curve which we are tracing. Then the first limit point is the first zero $s = s^* \geq 0$ of the real function

$$\phi(s) = (e^{\ell})^T Tx(s), \quad s \geq 0, \tag{6.6}$$

where the index ℓ, $n+1 \leq \ell \leq n+2$, was chosen such that (6.5) holds.

If there exists a total potential π and ϕ changes signs at $s = s^*$ then at that point of the curve one eigenvalue of $D_y\hat{F}(x(s^*))$ becomes positive and stability is certainly lost. Limit points where ϕ changes sign are called simple limit points. They may be characterized in terms of the behavior of the second derivative $D_{yy}F(x(s^*))$ but there is no need to discuss this here (see eg. [35]). We note also that zeroes of ϕ where the sign does not change are numerically ill-conditioned and cannot be detected easily. Accordingly, in practice, interest centers almost exclusively on simple limit points.

The continuation process produces a sequence of points $x^0, x^1, \ldots$ approximating our curve and we expect to encounter a first pair of points x^{k-1}, x^k where

$$\text{sign } (e^{\ell})^T Tx^{k-1} \neq \text{sign } (e^{\ell})^T Tx^k. \tag{6.7}$$

This signals the nearness of a simple limit point $x^* \in L(F,q)$ and the task (5.1) (ii) requires a closer computation of x^*. For this several algorithms are available (see the references given in [35]). A very simple but reliable approach is to apply a root-finder to the function ϕ. More precisely, we consider the secant line

$$z(t) = (1-t)x^{k-1} + tx^k, \quad 0 \leq t \leq 1 \tag{6.8}$$

between x^{k-1} and x^k and denote by i the index of the maximal component in modulus of $x^k - x^{k-1}$. By construction of our continuation process Newton's method applied to the system

$$\begin{pmatrix} \hat{F}x \\ (e^i)^T (x-z(t)) \end{pmatrix} = 0 \tag{6.9}$$

should converge to a point on the curve. Let $\tilde{x}(t)$ denote the accepted, final Newton-iterate corresponding to the starting point $z(t)$, $0 \leq t \leq 1$. Then the real function $\tilde{\phi}(t) = (e^\ell)^T T\tilde{x}(t)$, $0 \leq t \leq 1$, is well-defined and, by (6.7), it has different signs for $t = 0$ and $t = 1$. Hence, a standard root-finder can be applied to $\tilde{\phi}$. In particular, experience has shown that the code given in [21] works here very well.

In structural problems it is rarely sufficient to determine just one limit point. If, say, the structure is expected to buckle at that point then we are interested in determining the change of that buckling point with a variation of the parameters. The standard computational approach for this is to repeat the entire process, that is, to define a neighboring curve by changing the starting point and the augmenting equation (5.7) and

to determine on this new curve the first simple limit point. Obviously, this may be a very time-consuming method. Moreover, it is usually extremely difficult to construct neighboring curves which lead to a sequence of reasonably closely spaced limit points.

This raises the question about the feasibility of methods for a direct trace of a curve in the critical boundary $C(F) \cap R(F)$ through our first simple limit point x^*. Several such methods have been discussed in [35] and it would lead too far to go into details here. Briefly, one of the processes works with two augmented systems of the form

$$G_\ell(y,p,\bar{p}) = \begin{pmatrix} F(y,p) \\ (q^\ell)^T (p-\bar{p}) \end{pmatrix} = 0, \quad \ell = 1,2 \tag{6.10}$$

where $q^2 = q \in R^2$ and $q^1 \in R^2$ is orthogonal to q. In general, these systems define two families of curves on the equilibrium surface. Then the concept of the process is to start at the first limit point x^* with a trace of the curve through x^* defined by G_1. After one or at most a few continuation steps we switch to the curve defined by G_2 and determine on it the first simple limit point which represents the next point on our desired curve in the critical boundary (see Figure 7). A repetitive application of this concept requires attention to some technical details for which we refer to the cited article [35].

As an example of the application of this process we consider a clamped, thin, shallow, circular arch which has been used as a test case by various authors (see eg. [25], [26], [43]). Let U and W be the radial and axial displacements, R the arch-radius, A the cross-sectional area, h the thickness, and E Young's modulus. With the dimensionless displacements

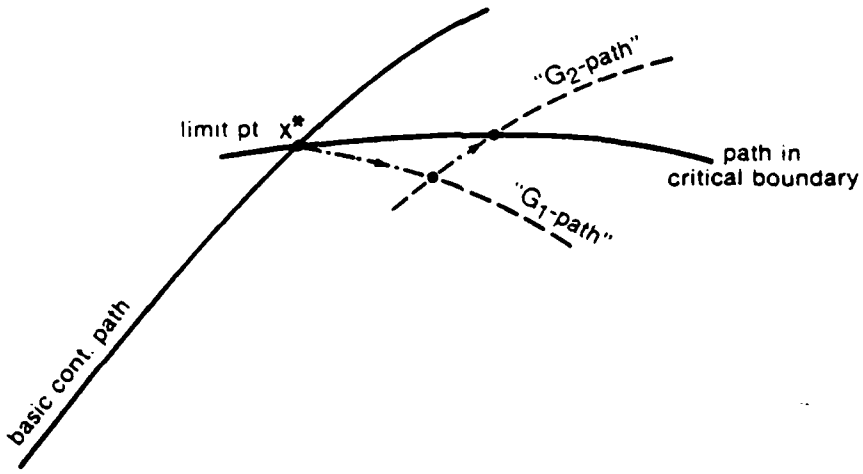

"G_2-path"
limit pt x^*
path in
critical boundary
"G_1-path"
basic cont. path

Figure 7

$u = U/n$, $w = W/h$, the total potential energy -- non-dimensionlized by dividing by $EAR(h/R)^2$ -- is given by

$$(6.11) \qquad \frac{1}{2}\int_{-\theta_o}^{\theta_o} [(\frac{dw}{d} - u) + \frac{1}{2}\frac{h}{R}(\frac{du}{d\theta})^2]^2 d\theta + \frac{\alpha_1}{2}\int_{-\theta_o}^{\theta_o} (\frac{d^2u}{d\theta^2})^2 d\theta - \alpha_2 \int_{-\theta_o}^{\theta_o} pu d\theta.$$

Here $p = p(\theta)$ is the dimensionless radial load and α_1, α_2 are dimensionless constants. Each end is assumed to be pinned, that is, we have the boundary conditions

$$(6.12) \qquad u(\pm\theta_o) = 0, \quad w(\pm\theta_o) = 0, \quad \frac{d^2u}{d\theta^2}(\pm\theta_o) = 0.$$

The finite element approximation introduced in [43] was applied. More specifically, we used a uniform mesh with eight elements, $\theta_o = 15^o$ and the constants $\alpha_1 = 3.8716 \times 10^{-6}$, $\alpha_2 = 8.2752 \times 10^{-2}$ corresponding to the data in [26]. The following four load functions $p = p(\mu,\nu)$ were chosen:

$$(6.13) \qquad \begin{aligned}
&(1) \quad p(\mu,\nu) = \begin{cases} \mu(1-\nu) + 4\mu\nu, & \text{for elements 4 and 5} \\ \mu(1-\nu), & \text{otherwise} \end{cases} \\
&(2) \quad p(\mu,\nu) = \begin{cases} \mu(1-\nu) + 4\mu\nu, & \text{for elements 2 and 3} \\ \mu(1-\nu), & \text{otherwise} \end{cases} \\
&(3) \quad p(\mu,\nu) = \begin{cases} \mu(1-\nu) + 8\mu\nu, & \text{for element 4} \\ \mu(1-\nu), & \text{otherwise} \end{cases} \\
&(4) \quad p(\mu,\nu) = \begin{cases} \mu(1-\nu) + 8\mu\nu, & \text{for element 2} \\ \mu(1-\nu), & \text{otherwise} \end{cases}
\end{aligned}$$

Hence, in all case the average load is μ.

If the boundary conditions are disregarded there are 45 displacement variables and two control parameters. In order to simplify the comparison for different boundary conditions, all variables were always numbered consecutively from 1 to 47. Thus, x_{21} denotes the radial displacement of the center point. Of course, once the boundary conditions are taken into account there remain only 39 displacement variables.

For each load function a basic curve on the equilibrium surface was defined by fixing a value of ν. More specifically, for the symmetric load (6.13) (1) the value $\nu = 0$ was used while in all other cases $\nu = 0.005$ was taken. These curves were followed numerically until a first limit point was detected. From this limit point the corresponding curve in the critical boundary was traced by the method sketched above. The results are shown in Figure 8. The curves in the critical boundary for the load functions (6.13) are labeled (1) through (4). They represent the location of the buckling points of the arch under the chosen load regimes. Note the considerable influence of the location of the load maxima for equal average loads. The curves (2), (3), (4) for the asymmetric loads emanate from the bifurcation point on the basic curve defined by $\nu = 0$. On these curves the structure buckles asymmetrically. On the other hand, for the symmetric load (1) the "critical curve" intersects at the limit point of the curve for $\nu = 0$, and represents points where the structure undergoes symmetric "snap through" buckling. It is easily seen that this is an instable situation which, in practice, is not realizable except under very controlled conditions.

The example shows that this approach can provide a wealth of information about the buckling behavior of a structure. Moreover, it turns out that the

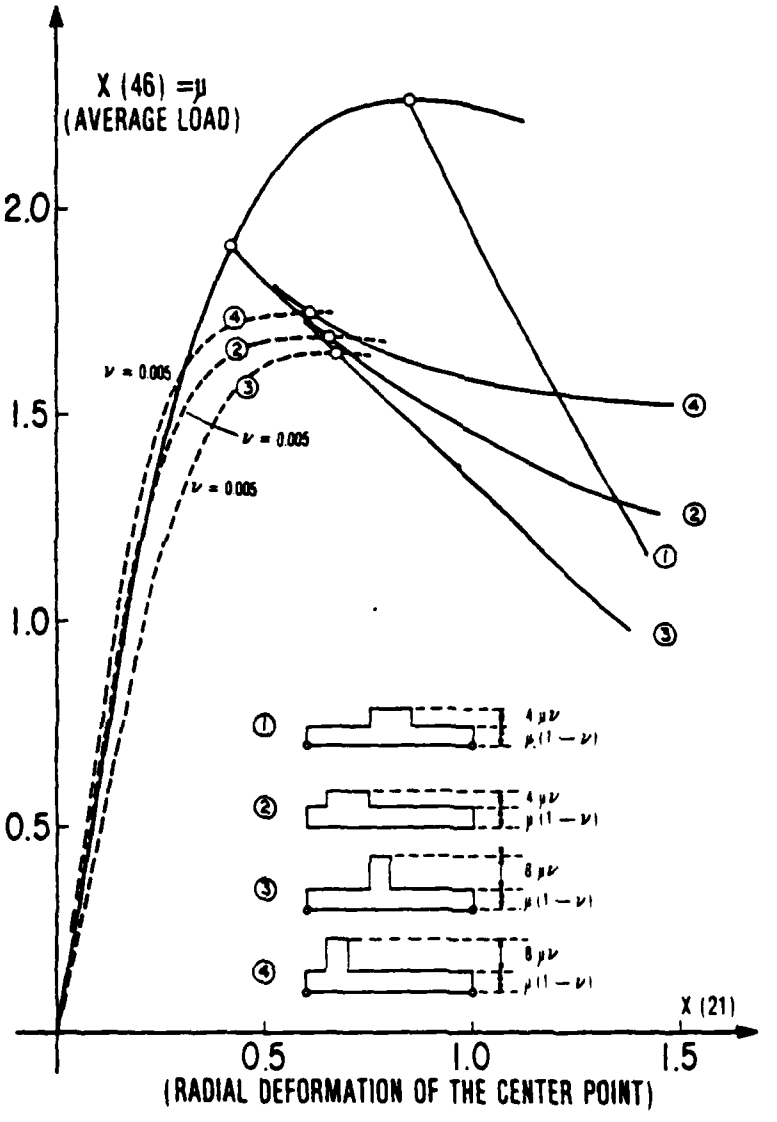

X (46) =μ
(AVERAGE LOAD)
2.0
1.5
1.0
0.5
ν = 0.005
ν = 0.005
ν = 0.005
X (21)
0.5
1.0
1.5
(RADIAL DEFORMATION OF THE CENTER POINT)

Figure 8

computational work for tracing one of the curves in the critical boundary is only mildly larger than that for tracing one of the basic curves. Thus, this type of analysis of a structure is certainly practically feasible.

7. A Posteriori Error Estimates and Adaptive Approaches

In Section 4 we introduced a particular finite element formulation for our model problem (2.10)/(2.11). This formulation is rather flexible and leaves the user considerable freedom in the choice of various data-items, including, in particular, the mesh (4.2), degree p of the elements and smoothness-index d of the solution. Now we face the following two tasks:

(i) Construct a posteriori error estimates which -- for given input-data-items -- characterize accurately and reliably the error caused by the finite-element solution as measured in some specified norm.

(ii) Design an adaptive procedure for selecting the controlling data-items defining the solution so as to achieve an acceptable solution with the prescribed accuracy.

These two tasks are closely tied together. *Without reliable* error estimators it is impossible to design effective adaptive procedures for solving the problem within the prescribed error-range. At the same time, it can be shown theoretically and practically that effective error estimators exhibit increasing accuracy whenever a proper adaptive approach is used.

The mentioned flexibility of our finite-element formulation for the model problem incorporates both the H-version and the P-version of the method. In other words, we may fix the degree of the elements and achieve the desired accuracy by refining the mesh, or, alternately, we may fix the mesh and increase the degree of the elements to meet the accuracy requirement. For some details and comparisons of these two versions see [14], [15], [41]. In this paper we shall concentrate on the H-version for which the approaches presented here are closely related to those analyzed for the linear case in [4], [5], [6], [9], [13].

We begin with a summary of the definition of the error-estimators for

the case $d = 1$ of C^0-element of degree p. The corresponding proofs will be given elsewhere.

For given parameter values λ, ν we denote by $u(s) = u(\lambda,\nu)(s)$ the exact solution of our problem. Moreover, for a specified mesh Δ of (4.2) and fixed degree p let $\tilde{u}(s) = u(\Delta,p,\lambda,\nu)(s)$, $0 \leq s \leq 1$, be the computed finite element solution. Then the error estimator is obtained as follows:

(i) On every sub-interval I_i, $i = 1,\ldots,n$, compute the finite element approximation

(7.1) $$w_i(s) = \sum_{i=0}^{p+1} z_{ij}\, \psi_j(\tau_i(s)), \quad \forall s \in I_i,$$

of the original problem subject to the boundary conditions

(7.2) $$w_i(s_\ell) = u(\Delta,p,\lambda,\nu)(s_\ell), \quad \ell = i-1,i.$$

(Here the notation of section 3 is used). Then, for each i, $1 \leq i \leq n$, define

(7.3) $$\sigma_i = [\int_{I_i} (w_i'(s) - \tilde{u}'(s))^2 ds]^{1/2}.$$

(ii) On every sub-interval I_i compute the residual

(7.4) $$r_i(s) = a(\tilde{u}'(s))' + \hat{b}(u(s),\lambda,\nu)$$

and the quantity

(7.5) $$\rho_i = c\, h_i^p \left[\int_{I_i} r_i(s)^2 ds\right]^{1/2}$$

where c is a fixed constant (see [13]). For small h_i we have $\rho_i << \sigma_i$ and hence this step (ii) can often be omitted.

(iii) With the quantities (7.3) and (7.5) the error indicator of the ith sub-interval is defined by

(7.6) $$\eta_i(\Delta) = [\sigma_i^2 + \rho_i^2]^{1/2}, \quad i = 1,\ldots,n.$$

(iv) Now the quantity

(7.7) $$\varepsilon(\Delta) = [\sum_{i=1}^{n} \eta_i(\Delta)^2]^{1/2}$$

is the desired error estimator of the computed solution.

The error estimator ε of (7.7) approximates the error

(7.8) $$||u(\lambda,\nu) - u(\Delta,p,\lambda,\nu)||_{H^1(I)} = [\int_0^1 (u'(s) - \tilde{u}'(s))^2 ds]^{1/2}$$

of the computed finite element solution in the H^1-norm. More precisely, under fairly general assumptions it is possible to prove that when the exact solution $u(\lambda,\nu)$ exists in a suitable neighborhood of the computed solution $u(\Delta,p,\lambda,\nu)$, then

(7.9) $$||u(\lambda,\nu) - u(\Delta,p,\lambda,\nu)||_{H^1(I)} = \varepsilon(\Delta)(1 + o(1)), \quad \text{as} \quad \varepsilon \to 0.$$

The term $o(1)$ depends on the exact solution. As we saw earlier the existence assumption is certainly needed if we want an estimate of the form (7.9). More generally, suppose that we trace a curve $\nu = \nu_0 = \text{const.}$ Then, for $\alpha > 1$ it can be shown that there exists an exact solution $u(\tilde{\lambda},\nu)$ such that

$$(7.10) \qquad ||u(\Delta,p,\lambda,\nu_0) - u(\tilde{\lambda},\nu_0)||_{H^1(I)} + |\lambda - \tilde{\lambda}|^{\alpha} = \varepsilon(\Delta)(1 + o(1)) \quad \text{as} \quad \varepsilon \to 0.$$

For a solution curve defined by $\nu = \nu_0$ the situation corresponding to the two estimates (7.9) and (7.10) was schematically indicated in Figure 6.

The error estimator (7.7) is composed of the error indicators (7.6) for the individual sub-intervals of the given mesh. For the linear case we have shown in [9] that an (asymptotically) optimal mesh is achieved if we equilibrate the error indicators, that is, if we construct a mesh for which these indicators are equal. More specifically, we call a sequence Δ_k of meshes equilibrated if

$$(7.11) \qquad \frac{\max \eta_i(\Delta_k)}{\min \eta_i(\Delta_k)} \le K, \quad \forall k,$$

with a constant K that is independent of the particular mesh Δ_k. These concepts provide the basis for an adaptive mesh-refinement procedure which generates a sequence of meshes that satisfy (7.11) and for which the error indicators are approximately equal, (see [8], [10], [11], [12], [46]). As noted before, the equilibration of the meshes is expected also to increase the effectivity of the error estimator (see [12], [13]).

Suppose again that we trace a curve with $\nu = \nu_0 = \text{const.}$ As we observed in Sections 3 and 6 the position of a turning or limit point on

this curve frequently has considerable practical interest. More specifically, we are interested in estimating the error $|\lambda_T - \tilde{\lambda}_T|$ indicated in Figure 6. Let $\tilde{u}_T = u(\Delta,p,\tilde{\lambda}_T,\nu_0)$ be the computed solution at the computed critical point. Then an a posteriori estimate of $|\lambda_T - \tilde{\lambda}_T|$ may be obtained by the following procedure:

(i) Compute the indicators σ_i, $i = 1,\ldots,n$, of (7.3) and the quantity

$$\sigma = [\sum_{i=1}^{n} \sigma_i^2]^{1/2}. \tag{7.12}$$

(ii) Let ϕ be the function corresponding to the tangent vector (5.12) at the computed critical point and form

$$\kappa = 0.05 \frac{\sigma}{\|\phi\|_{H^1(I)}}. \tag{7.13}$$

(iii) On every sub-interval I_i, $i = 1,\ldots,n$, compute the finite element approximation $\bar{w}_i$ of the form (7.1) subject to the boundary conditions

$$\bar{w}_i(s_\ell) = \tilde{u}_T(s_\ell) + \kappa\, \phi(s_\ell), \quad \ell = i-1,i. \tag{7.14}$$

(iv) Form the quantities

$$\tau = \sum_{i=1}^{n} \int_{I_i} [-a(\tilde{u}_T')(\bar{w}_i' - w_i') + \hat{b}(\tilde{u}_T,\lambda_T,\nu_0)(\bar{w}_i - w_i)]ds \tag{7.15}$$

and

$$\zeta = \int_I \frac{d}{d\lambda} \hat{b}(\tilde{u}_T,\lambda_T,\nu_0)\kappa\, \phi\, ds \tag{7.16}$$

where the last quantity represents a linearization of the non-linear function of λ

$$\int_I [\hat{b}(\tilde{u}_T,\lambda,\nu_0) - \hat{b}(\tilde{u}_T,\lambda_T,\nu_0)]\kappa\ \phi\ ds. \tag{7.17}$$

(v) Then the quotient

$$\theta = \left|\frac{\tau}{\zeta}\right| \tag{7.18}$$

is the desired estimator.

Analogously as in the case of the estimator of the solution it is possible to show that

$$|\lambda_T - \tilde{\lambda}_T| = \theta(1 + o(1)), \quad \text{as} \quad \varepsilon \to 0 \tag{7.19}$$

where the term $o(1)$ depends on the exact solution.

8. *Numerical Experiments*

The theory outlined in the previous section raises several questions, as, for instance, the following:

(i) In practice, how effective are the error estimators for the computed solution and the critical values, especially in view of the fact that the theory of these estimators is only asymptotic in nature?

(ii) Is the quality of the estimators influenced by the type of nonlinearity, as exemplified for instance by the three problems (2.14)?

(iii) How effective is the adaptive procedure based on the equilibration of the error indicators, and, in particular, its use in combination with the continuation process of Section 5?

In order to provide some answers to these and related questions we present here some computational results obtained with our model problem.

We consider first the strongly nonlinear rod (2.14) (i) and, in particular, the general shape of its equilibrium surface. Figure 9 corresponds in essence to Figure 4 for the simple framework of Section 3. More specifically, uniform meshes with eight subintervals and cubic C^0-elements were used, and Figure 9 shows several solution curves for ν = const. plotted in a coordinate system with λ as the abscissa and the norm

$$||\tilde{u}||_{L_\infty} = \max_{0\le x\le 1} |\tilde{u}(x)| \tag{8.1}$$

as ordinate. Note that in contrast to the case of the framework of Figure 2 the solutions now have two turning points. The first bifurcation point occurs approximately at $\lambda \doteq 2.8954$, $\nu = 0$, and the second one at $\lambda \doteq 6.5691$, $\nu = 0$. For the linearized equation the corresponding first two eigenvalues are $\lambda \doteq 2.8954$ and $\lambda \doteq 6.1448$.

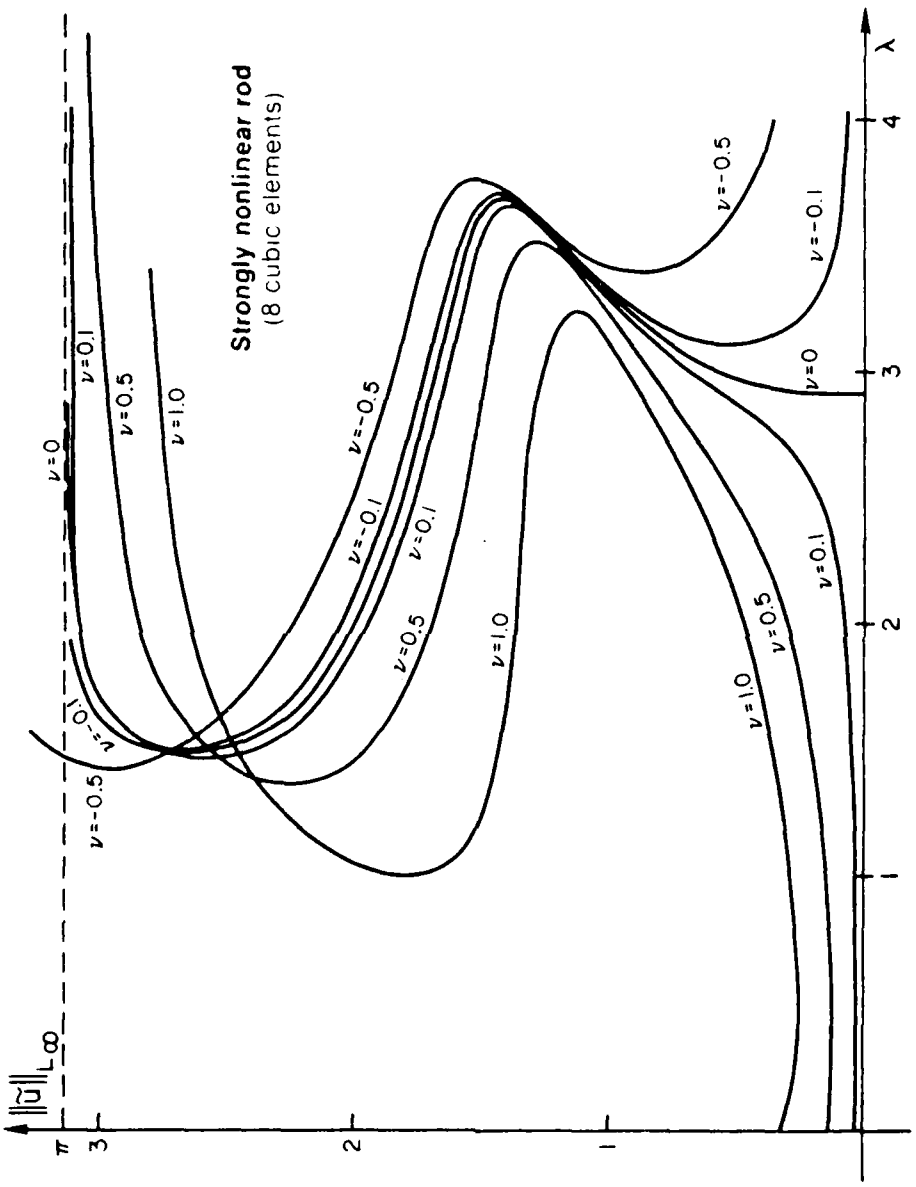

Strongly nonlinear rod
(8 cubic elements)
$\|\tilde{u}\|_{L_\infty}$
λ
π
3
2
1
1
2
3
4
$\nu=0$
$\nu=0.1$
$\nu=0.5$
$\nu=1.0$
$\nu=-0.1$
$\nu=-0.5$
$\nu=-0.5$
$\nu=-0.1$
$\nu=0.1$
$\nu=0.5$
$\nu=1.0$
$\nu=1.0$
$\nu=0.5$
$\nu=0.1$
$\nu=0$
$\nu=-0.1$
$\nu=-0.5$

Figure 9

Figure 10 gives a corresponding picture of the solution curves with the moment balance $M(0) - M(1)$ (see (2.16)) as ordinate in place of (8.1). Figures 11 and 12 show the shape of the rod for $\nu = 0.1$ and variable λ and for $\lambda = 3.0$ and variable ν, respectively.

In order to assess the effectivity of the error estimation, we used as "nearly exact" solution the computed solution with a uniform mesh of 10 subintervals and quintic C^0-elements. We consider the curve $\nu = 0.1$ and on it the five target points $\lambda = 1.0, 2.5, 3.5, 2.5, 1.6$. Table 1 gives the computed moment balances $M(0) - M(1)$ at these five points for a uniform mesh with eight subintervals and C^0-elements of different order

λ \ p	1	3	5
1.0	0.1203	0.1206	0.1206
2.5	0.5362	0.5558	0.5558
3.5	2.595	2.636	2.636
2.5	2.787	2.500	2.500
1.6	2.845	2.830	2.830

Table 1

Table 2 shows the computed error estimators and the "nearly exact" error norm in percent of the norm of the solution. In these and all other computations in this section the (small) contribution of the ρ_i-terms of (7.5) was neglected. The table shows clearly the high quality of the estimators. In the cases $p = 4$, $\lambda = 1.0$ and $\lambda = 2.5$ left blank in the table the error is essentially zero within machine accuracy.

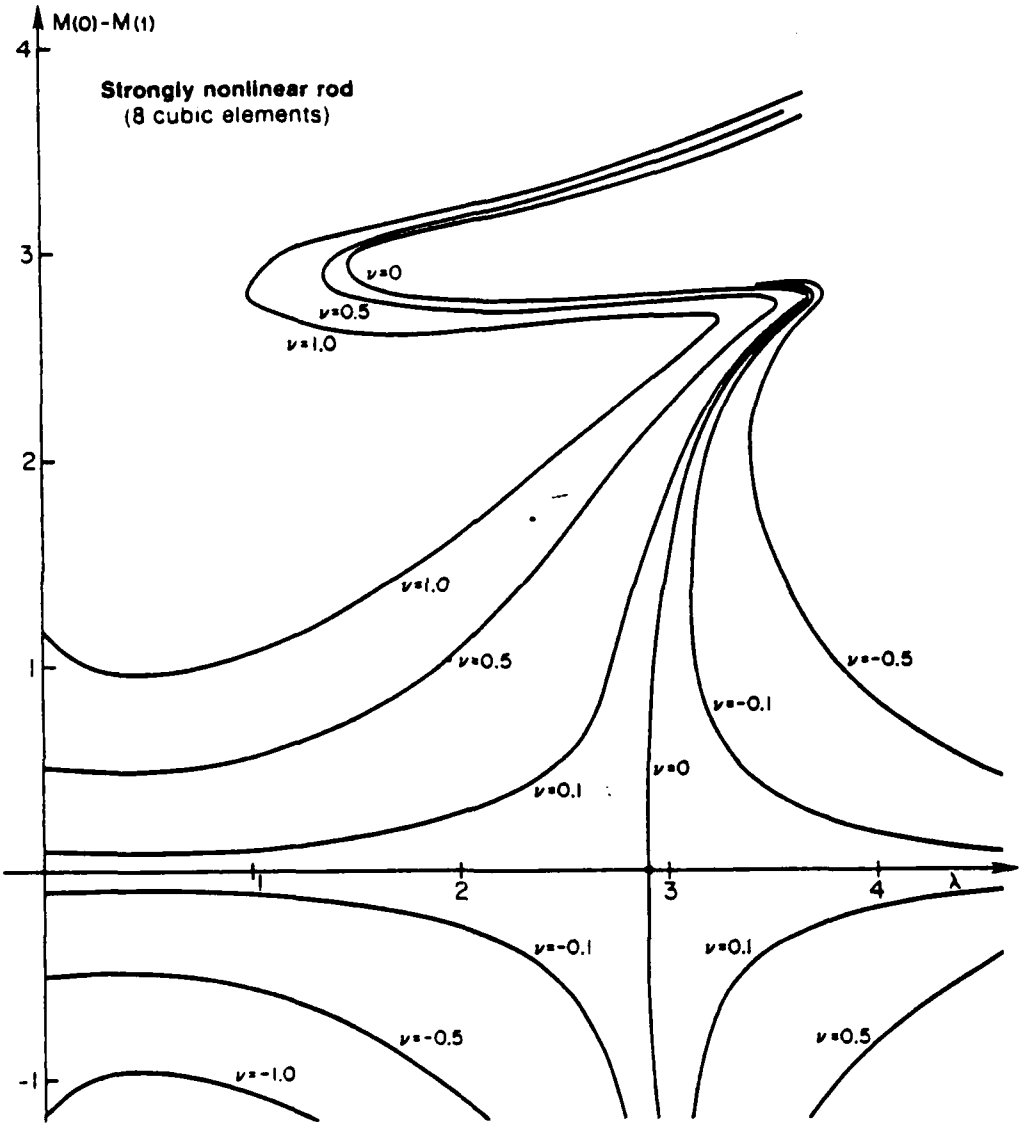

M(0)-M(1)
Strongly nonlinear rod
(8 cubic elements)
4
3
2
1
-1
1
2
3
4
λ
ν=0
ν=0.5
ν=1.0
ν=1.0
ν=0.5
ν=0.1
ν=0
ν=-0.1
ν=-0.5
ν=-0.1
ν=0.1
ν=-0.5
ν=0.5
ν=-1.0

Figure 10

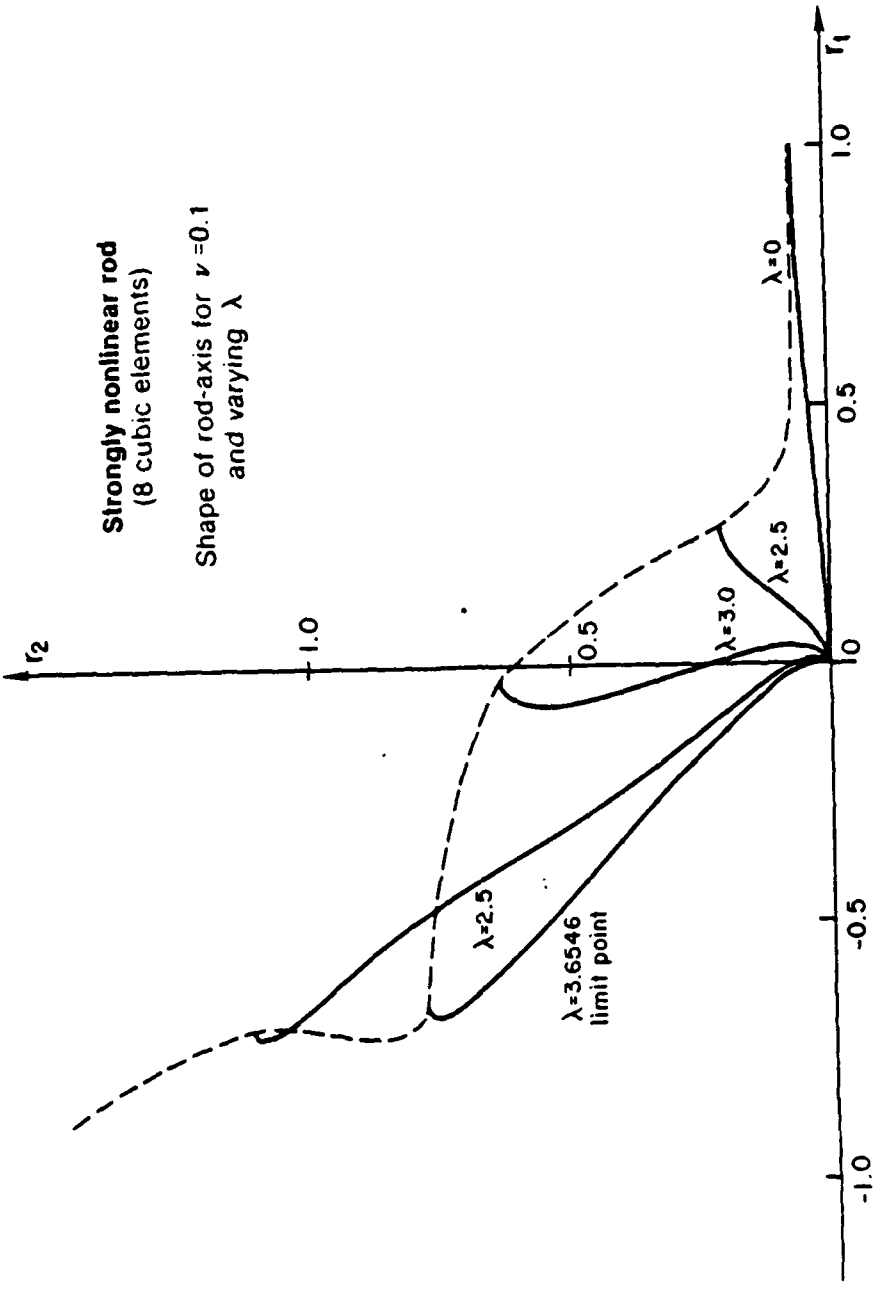

Strongly nonlinear rod
(8 cubic elements)
Shape of rod-axis for ν=0.1
and varying λ
r2
1.0
0.5
0
-0.5
-1.0
r1
1.0
0.5
λ=0
λ=2.5
λ=3.0
λ=2.5
λ=3.6546
limit point

Figure 11

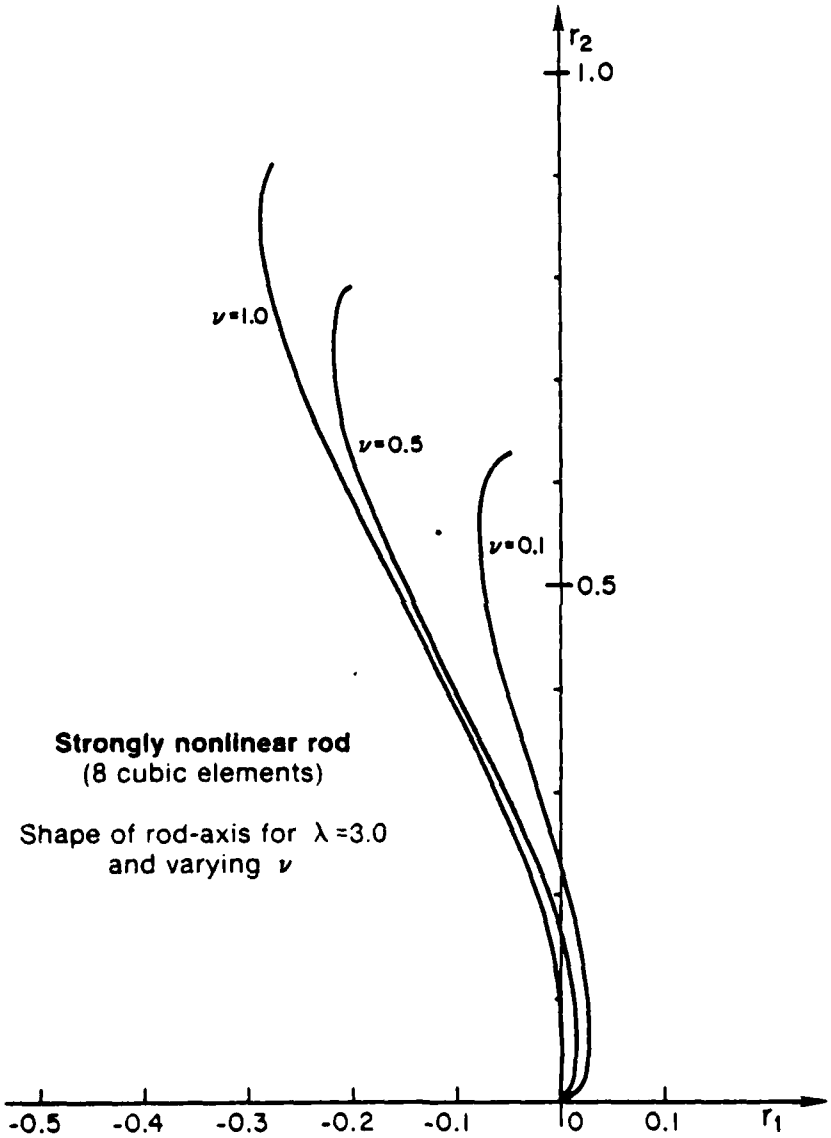

r_2
1.0
0.5
ν=1.0
ν=0.5
ν=0.1
Strongly nonlinear rod
(8 cubic elements)
Shape of rod-axis for λ =3.0
and varying ν
-0.5
-0.4
-0.3
-0.2
-0.1
0
0.1
r_1

Figure 12

	λ \ p	1	2	3	4
Error estim.	1.0	12.15%	0.1137%	0.004089%	
Error norm		12.13%	0.1138%	0.004089%	
Error estim.	2.5	11.55%	0.4306%	0.01552%	
Error norm		12.04%	0.4310%	0.01552%	
Error estim.	3.5	21.58%	1.597%	1.123%	0.5011%
Error norm		23.91%	1.900%	1.391%	0.5389%
Error estim.	2.5	21.53%	1.708%	1.003%	0.5013%
Error norm		21.88%	1.865%	1.285%	0.5455%
Error estim.	1.6	19.45%	0.7632%	0.8123%	0.3171%
Error norm		19.14%	1.135%	0.9576%	0.3312%

Table 2

As Figure 10 shows there are two turning points along the curve for $\nu = 0.1$. With a mesh of eight equal, cubic C^0-elements these turning points occur at

$$\lambda_{T_1} \doteq 3.6546, \quad \lambda_{T_2} \doteq 1.46749. \tag{8.2}$$

In order to assess the effectivity of our error-estimators of these critical value we used for the comparison five and ten equal, quadratic C^0-elements, respectively.

Table 3 gives the result of these computations. Again the error estimators perform very well. The table also shows that convergence of the critical λ-values is faster than that of the corresponding solution errors. In fact, asymptotically it is expected that the critical values converge with the square of the solution errors in the H^1-norm.

Limit Point	No. of Interv.	Est. of Solution Error	Limit Point Value	Limit Point Error	Limit Point Estim.
λ_{T_1}	5	13.11%	3.699385	0.04472	0.05294
	10	3.18%	3.657982	0.00317	0.00399
λ_{T_2}	5	7.205%	1.470290	0.002793	0.002843
	10	1.613%	1.467897	0.00040	0.000357

Table 3

In preparation for the use of the error estimators for an adaptive construction of optimal meshes, we show in Table 4 the error indicators of the solution at $\lambda = 3.5$, $\nu = 0.1$ for uniform C^0-elements of varying order. Evidently for lower p the mesh is better equilibrated than for

i \ p	1	2	3	4	5
1	0.4085	0.2849(-1)	0.2525(-1)	0.1133(-1)	0.2058(-3)
2	0.1789	0.2149(-1)	0.2816(-2)	0.3089(-3)	0.2637(-4)
3	0.9545	0.5290(-2)	0.5770(-3)	0.4769(-4)	0.4212(-5)
4	0.7281	0.1173(-2)	0.2176(-3)	0.7614(-5)	0.9042(-6)
5	0.7281	0.1173(-2)	0.2176(-3)	0.7614(-5)	0.9060(-6)
6	0.9545	0.5290(-2)	0.5770(-3)	0.4769(-4)	0.4212(-5)
7	0.1789	0.2149(-1)	0.2816(-2)	0.3089(-3)	0.2637(-4)
8	0.4085	0.2849(-1)	0.2525(-1)	0.1133(-1)	0.2058(-3)
$\frac{\max \eta_i}{\min \eta_i}$	5.34	24.3	116.	1488.	228.

Table 4

higher values. (In the case $p = 5$ in Table 4 the very small error indicators for the intervals with index $i = 4$ and $i = 5$ are distortet by round off which causes their values to be too large.) The results of Table 4 indicate that an adaptive procedure for the construction of the mesh should be effective.

As Table 2 shows the error may vary considerably along any one solution curve. Suppose that we wish to compute the curve for $\nu = 0.1$ with an accuracy of 2%. For this we begin with a mesh that meets this requirement; in our case, a uniform mesh of two quadratic C^0-elements. Then during the

continuation process we monitor the error estimators and, at any computed point where the estimate exceeds 2%, we divide the element(s) with the maximal error indicator. Table 5 describes the result of this adaptive procedure for the curve $\nu = 0.1$. Quadratic elements were used throughout.

λ	Error Indicator min	Error Indicator max	Rel. Error Estimate %	# of Intervals	Comments
0	.433(-5)	.433(-5)	0.0106	2	
0.5	.299(-3)	.299(-3)	.727	2	
1.5	.201(-2)	.201(-2)	2.81	2	Refine
	.183(-3)	.521(-3)	0.770	4	
2.49	.129(-2)	.421(-2)	1.68	4	
2.79	.547(-3)	.122(-1)	1.96	4	
3.37	.156(-1)	.115(-1)	5.96	4	Refine
	.482(-2)	.170(-1)	1.19	6	
3.55	.164(-1)	.491(-1)	2.38	6	Refine
	.137(-1)	.228(-1)	1.51	8	
3.63	.139(-1)	.249(-1)	1.45	8	
3.66	.157(-1)	.260(-1)	1.36	8	Limit Point
3.63	.216(-2)	.266(-1)	1.33	8	
3.44	.168(-2)	.272(-1)	1.27	8	
3.16	.909(-2)	.272(-1)	1.22	8	
2.75	.613(-2)	.263(-1)	1.21	8	
1.84	.107(-1)	.482(-1)	1.58	8	
1.54	.148(-1)	.644(-1)	1.69	8	
1.47	.206(-1)	.752(-1)	1.94	8	
1.47	.207(-1)	.943(-1)	2.08	8	Limit Point

Table 5

The run was terminated at the second limit point otherwise, at that point a refinement would have taken place.

In comparison, the uniform mesh with the least number of quadratic C^0-elements for which the error along the same curve does not exceed 2% is the mesh with 12 elements. Table 6 gives the corresponding results for this uniform mesh.

λ	Error Indicators min	Error Indicators max	Rel. Error Estimate %	Comments
0	.271(-7)	.738(-7)	.310(-3)	
0.5	.570(-6)	.615(-5)	.231(-1)	
1.5	.393(-5)	.335(-4)	.874(-1)	
2.47	.258(-4)	.303(-3)	.189	
2.67	.283(-4)	.638(-3)	.214	
2.83	.202(-4)	.156(-2)	.257	
3.02	.154(-3)	.508(-2)	.446	
3.28	.256(-3)	.144(-1)	.925	
3.33	.262(-3)	.236(-1)	1.09	
3.66	.247(-3)	.378(-1)	1.65	Limit Point
2.99	.543(-4)	.262(-1)	.976	
1.90	.110(-2)	.217(-1)	.809	
1.57	.108(-2)	.223(-1)	.742	
1.48	.737(-3)	.286(-1)	.761	
1.47	.598(-3)	.983(-1)	2.07	Limit Point

Table 6

In order to compare the results of Tables 5 and 6 we define the work per continuation step as

$$f_1 m + f_2 m^{2.5} \text{ units,} \tag{8.3}$$

where f_1 and f_2 are the number of function and derivative evaluations, respectively, and m is the number of degrees of freedom (see Section 4). For the computation with the adaptive procedure, as shown in Table 5, the total work was 45,672 units while for the run with the uniform mesh, given in Table 6, it was 143,383 units. Thus, the adaptive process requires only 31.9% of the work for the fixed uniform mesh. At the same time, a maximum of eight elements were needed during the adaptive process while the fixed uniform mesh with the same accuracy required twelve elements. These results are typical for problems of this type. The error usually varies considerably along any solution curve and it is practically impossible to design a fixed uniform, or, for that matter, non-uniform mesh for which it can be guaranteed that the error remains below a prescribed accuracy. Only an adaptive process which modifies the mesh in response to the computed error estimates can accomplish this task. In our case we used only a mesh-refinement strategy, but, obviously, it is also possible to incorporate de-refinement when the estimates fall below a lower threshold.

It may be of interest to analyze why the fixed uniform meshes do not perform well in the higher parts of our solution curve for $\nu = 0.1$. For this we use a uniform mesh of eight, cubic C^0-elements and give in Table 7a the error indicators for the first three intervals at various target points along the curve $\nu = 0.1$. In addition, Table 7b shows the actual values of the function u and its derivative u' at the corresponding nodal points and the midpoints of these intervals. Clearly, in the upper parts of the

curve boundary layers develop near the ends of the rod and this is reflected by the error indicators.

i \ λ	2.5	3.5	2.5	λ = 1.4675 Limit Point
1	0.1703(-4)	0.2524(-1)	0.3200(-1)	0.2778(-1)
2	0.2358(-4)	0.2816(-2)	0.3924(-2)	0.5590(-2)
3	0.2218(-4)	0.5770(-3)	0.7445(-3)	0.2720(-2)

Table 7a

	λ = 2.5		λ = 3.5		λ = 2.5		λ = 1.4675	
x	u	u'	u	u'	u	u'	u	u'
0	0	0.5707	0	2.730	0	11.17	0	20.74
0.0625	0.03482	0.5409	0.1648	2.515	0.5833	7.639	1.048	13.20
0.1250	0.06726	0.4946	0.3106	2.133	0.9728	5.013	1.702	8.126
0.1875	0.09632	0.4335	0.4304	1.703	1.241	3.644	2.091	4.597
0.2500	0.1212	0.3606	0.5240	1.299	1.438	2.714	2.311	2.636
0.3125	0.1412	0.2782	0.5937	0.9358	1.585	1.989	2.442	1.628
0.3750	0.1558	0.1892	0.6417	0.6044	1.687	1.297	2.521	0.9527

Table 7b

So far we discussed only computations with C^0-elements. In order to assess the influence on the error of different degrees of smoothness we selected three cases with the same number of degrees of freedom (4.5), namely $m = 10$. The corresponding results for the solution curve defined by $\nu = 0.1$ in the range $0 \leq \lambda \leq 1.467$ up to the second limit point are given in Table 8:

Smoothness	Degree	No. of Elements	Max. Error in Range	Work Units
C^0	1	11	23.81%	22,243
C^1	3	5	3.939%	20,236
C^2	5	3	0.9976%	20,919

Table 8

The effectivity of the P-version inherent in these results corresponds to that in the linear case discussed in [2], [15].

All results given so far in this section concerned the strongly non-linear rod (2.14) (i). The results are essentially analogous to that for the mildly non-linear case. However, the turning points now occur for larger values of λ. This can be seen from Figure 13 where we give the first bifurcating solution for $\nu = 0.0$ in the three cases (2.14). Ten equal, quintic C^0-elements were used in each case. The effectivity of the error estimators also remains the same in the two cases a comparison of Tables 2 and 9 shows.

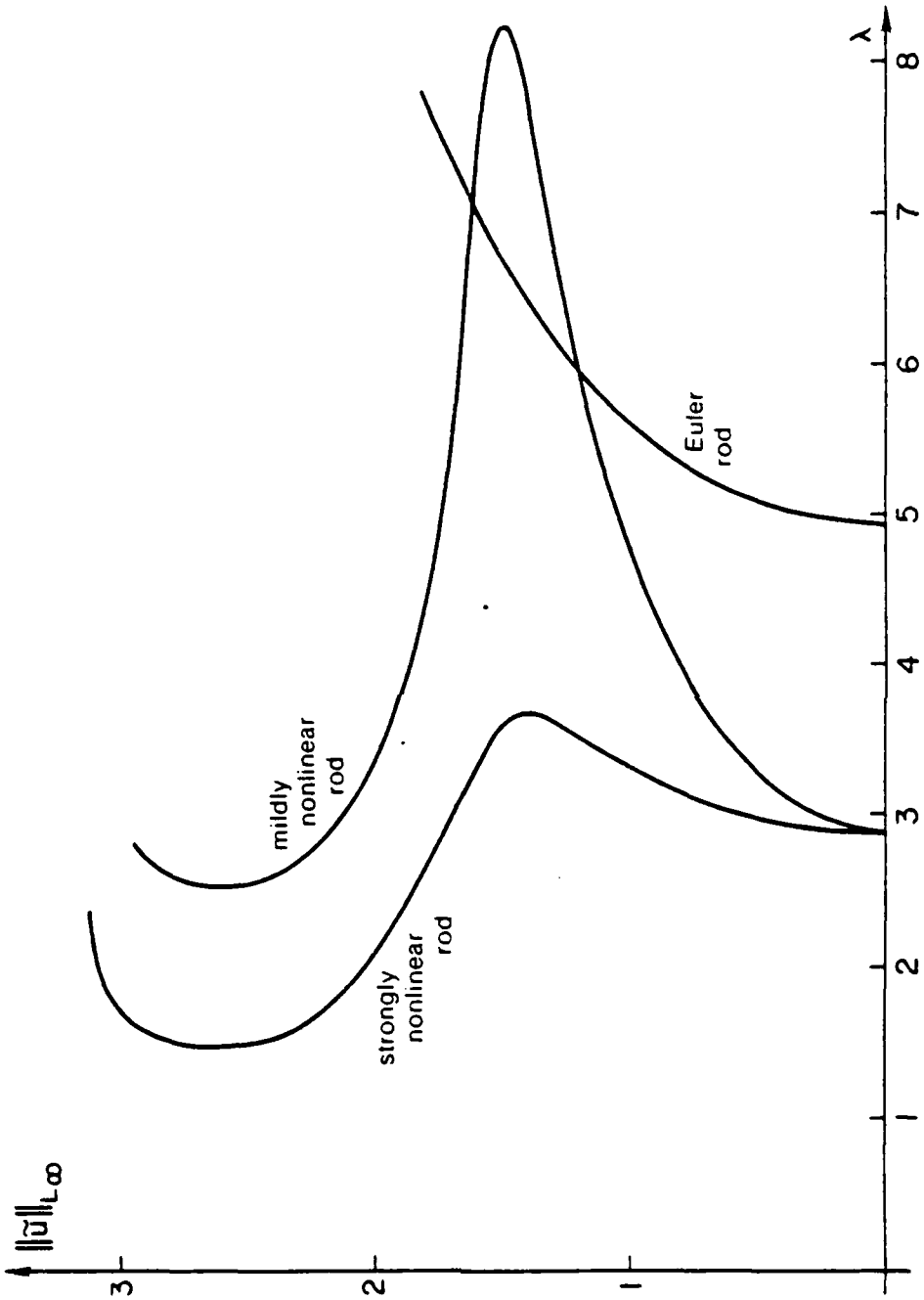

$\|\tilde{u}\|_{L^\infty}$
3
2
1
1
2
3
4
5
6
7
8
λ
mildly nonlinear rod
strongly nonlinear rod
Euler rod

Figure 13

	λ \ b	1	2	3	4
Error estim.	3.0	11.45%	.5184%	.01964%	.6520(-3)
Error norm		11.61%	.5187%	.01965%	.6535(-3)
Error estim.	5.0	11.87%	.7603%	.1023%	.3493(-2)
Error norm		11.92%	.7672%	.1024%	.2906(-2)
Error estim.	7.0	12.73%	.8964%	.2426%	.3202(-1)
Error norm		12.83%	.9296%	.2447%	.3202(-1)
Error estim.	5.0	14.49%	1.790%	.1260%	.8446(-1)
Error norm		12.26%	1.797%	.1518%	.8448(-1)
Error estim.	3.0	13.67%	1.166%	.2288%	.1595(-1)
Error norm		12.81%	1.189%	.2296%	.1850(-1)

Table 9

In summary the computational results of this section provide support for the following claims:

a) The a posteriori estimators of the errors of the solution and the critical values are very reliable for different types of nonlinearities and all degrees of elements.

b) Higher order elements perform in general better than the lower order ones.

c) The adaptive procedures for modifying the mesh during the continuation process is very effective and significantly increases the effectivity of the computations.

9. Spurious Solutions

As mentioned in the Introduction, for nonlinear problems of the type considered here the equilibrium surface of the discretized equations often has a different number of connected components than that of the original equations. The components which do not approximate reasonably well the exact solutions in any of their parts have been called spurious, or numerically irrelevant solutions. They have been observed by many authors (see eg. [20], [28] and the references given there).

The phenomenon is easily observed in the case of the classical Euler rod (2.14) (iii). For $\nu = 0$ it can be shown that for $\lambda < 0$ there exists no other solution than $u \equiv 0$. This is physically plausible since under pure tension the rod should remain straight. In fact, Figure 13 showed that the first non-zero solution bifurcates from the trivial solution at $\lambda \doteq 4.940$, and hence that we should expect non-zero solutions for $\nu = 0$ only when λ exceeds that value. Nevertheless, the finite element method does produce non-zero solutions for certain negative values of λ. Figure 14 shows this for the cases of uniform meshes with two, four, and six linear C^0-elements, respectively. The limit points with respect to λ for these three paths as well as for that one obtained with eight linear elements, are given in Table 10.

Number of Elements	λ_T
2	69.629
4	47.404
6	64.246
8	110.907

Table 10

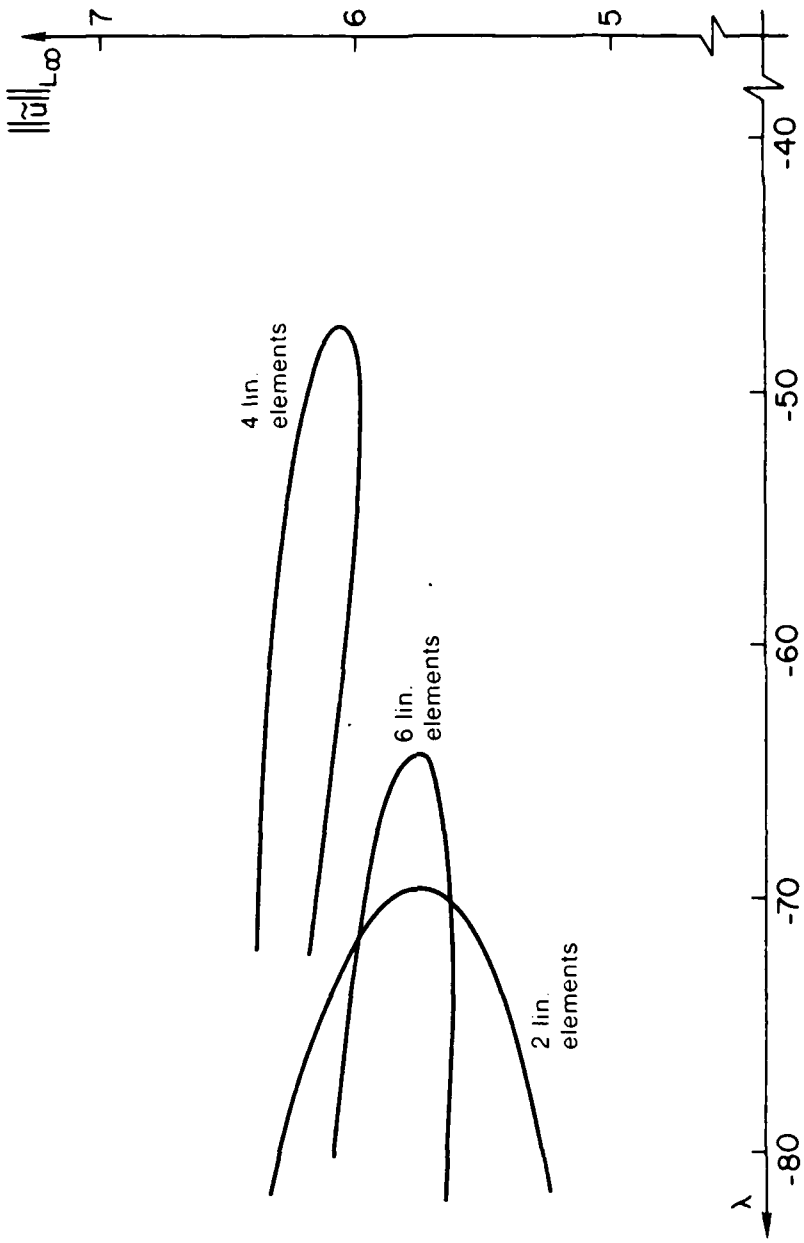

$\|\tilde{u}\|_{L^\infty}$
7
6
5
4 lin. elements
6 lin. elements
2 lin. elements
-40
-50
-60
-70
-80
λ

Figure 14

In this simple case it is easy to distinguish between the correct solutions for $\lambda > 0$ and the spurious ones for $\lambda < 0$. In general, however, the situation is more complicated and it is not readily clear how to distinguish between the correct and the spurious solutions. Under fairly general conditions the finite element solutions converge to the solutions of the infinite dimensional problem. Hence, in general, the spurious solutions have to disappear toward infinity when the mesh is refined. But this need not happen monotonically as Table 10 shows.

In line with our overall philosophy, we are interested in computing an approximate solution for which the a posteriori estimate indicates a deviation of at most 10-15% from an exact solution. In other words, we will accept only those segments of any computed solution path for which the estimate does not exceed, say, 12%. The spurious solutions should carry a much larger error and hence could be rejected on that basis. This raises the question whether our error estimators will succeed in identifying properly such large errors even though they were justified theoretically only asymptotically for small errors.

The following tables show that this is indeed the case. More specifically, Table 11 gives some error estimates for the spurious solution path obtained with a uniform mesh of eight linear C^0-elements.

λ	$\|\|u\|\|_{H^1}$	Error Estimate
-128.502	15.815	26.9%
-125.507	15.688	27.0%
-122.513	15.553	27.2%
-119.522	15.409	27.5%
-116.534	15.250	27.8%
-113.554	15.064	28.4%
-110.907*	14.758	30.4%
-112.192	14.632	32.5%
-115.796	14.581	35.0%
-119.599	14.585	36.8%
-123.491	14.611	38.4%
-127.417	14.651	39.9%
-131.359	14.699	41.3%

* Limit Point

Table 11

Since $u \equiv 0$ is the only exact solution for negative λ the value of $||u||_{H^1}$ represents here the actual error in the H^1-norm. In other words, the actual error is 100%, and, not surprisingly, the error estimators only have an efficiency index of about 0.27-0.41. Yet in line with the earlier discussion any solution path such as this one with errors exceeding the 12% threshold should be rejected.

On the other hand, Table 12 gives the error estimates for the "correct" solution path bifurcating from the positive λ-axis at about $\lambda \doteq 5$. Again a uniform mesh with eight linear elements was used.

λ	Error Estimates
5.1390	11.43%
5.3264	11.43%
5.5972	11.41%
5.9553	11.41%
6.3551	11.41%
6.7940	11.42%
7.3081	11.45%
7.8960	11.49%
8.5529	11.54%
9.2715	11.62%
10.042	11.71%
10.856	11.82%
11.704	11.95%
12.579	12.09%

Table 12

If we allow a maximum error of 12% then the entire path segment up to $\lambda \doteq 12.579$ should be accepted. On the other hand, if the maximum error is 10% then this path must be rejected and higher order elements are needed. Hence, the question of accepting or rejecting a path segment rests solely on our error estimate.

Part II: Outlook on More General Linear Problems

The results discussed in Part I above are as yet still rather restrictive in their applicability. Clearly, we should hope to be able to extend them to boundary value problems in several space dimensions of the type occurring in nonlinear structural analysis. But there are other possible extensions as well. In this part we illustrate some of the desired extension by means of results already obtained for the linear case.

More specifically, we begin with some computational examples of the use of our adaptive finite element solver FEARS. This system is based on our results cited in the Introduction about error estimators and adaptive procedures for two-dimensional elliptic problems. Then we turn to a simple parabolic model problem to show the possibility of applying these ideas also in the case of certain time-dependent problems.

As indicated before the material in this Part II is strictly intended only to indicate the type of results we would hope to extend also to the nonlinear case. For more details on these linear problems we refer to the cited articles.

10. An Adaptive Finite Element Solver

For linear elliptic problems in two-space dimensions the theory of a posteriori error estimates has been advanced to a point (see eg. [4], [5], [11]) that it became possible to implement a general finite element solver which incorporates the ideas outlined here, (see eg. [8], [36], [46]). More specifically, the FEARS system [Finite Element Adaptive Research Solver] was developed at the University of Maryland and is available in FORTRAN for use on Univac 1100 series computers and on the CDC-7600. It has the following principal features:

(i) FEARS constitutes an applications-independent finite element solver for a certain class of two-dimensional, linear, elliptic boundary value problems defined by a weak mathematical formulation. At present curvilinear elements of degree 1 are used.

(10.1) (ii) Adaptive approaches are employed extensively. The a posteriori estimates developed by us (loc.cit) are used to control the adaptive processes and to provide a solution with a near optimal error.

(iii) In the systems design, advantage was taken of the natural modularity of the finite element method (see [45], [46]).

Some results obtained with the system were described in [12]. We present here some further results which illustrate the effectivity of the adaptive approach.

Example 1: We consider the elasticity problem for an isotropic homogeneous material with Young's modulus $E = 1$ and Poisson's ratio $\nu = 0.3$ on the domain shown in Figure 15 where also the boundary conditions are specified. The energy norm is used for the error estimation. The basic mesh consisted of eight equal, bilinear element and Figure 16 a-j provides a sequence of

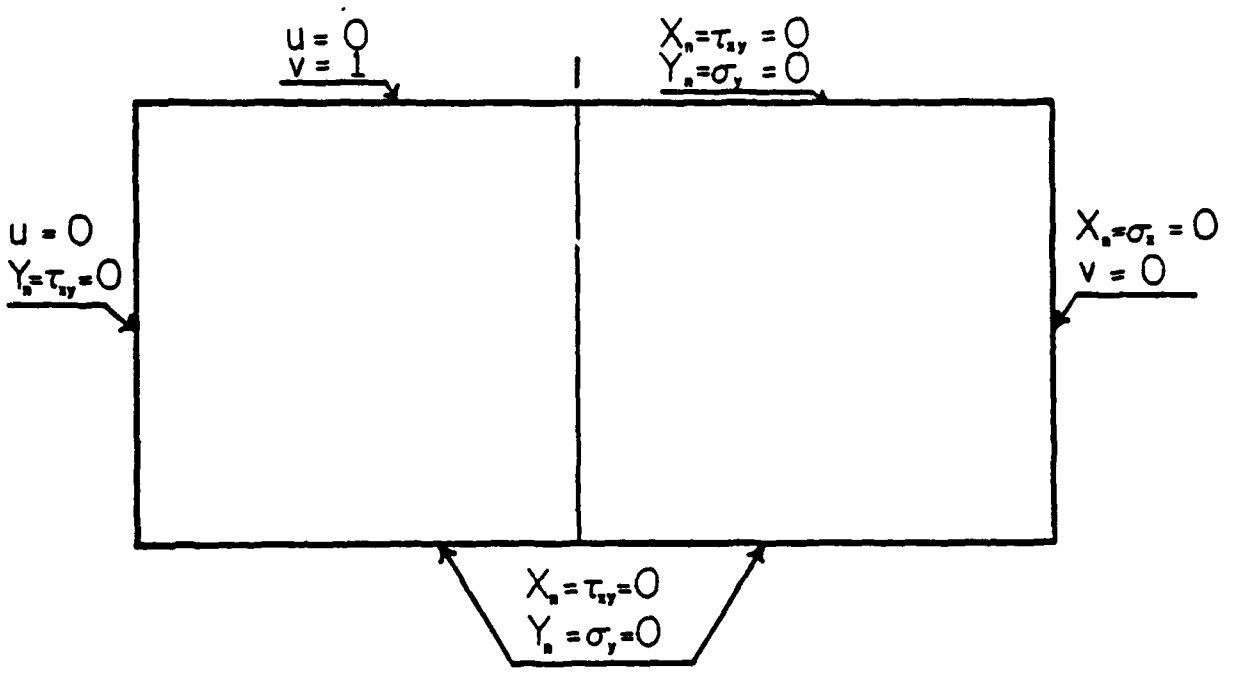
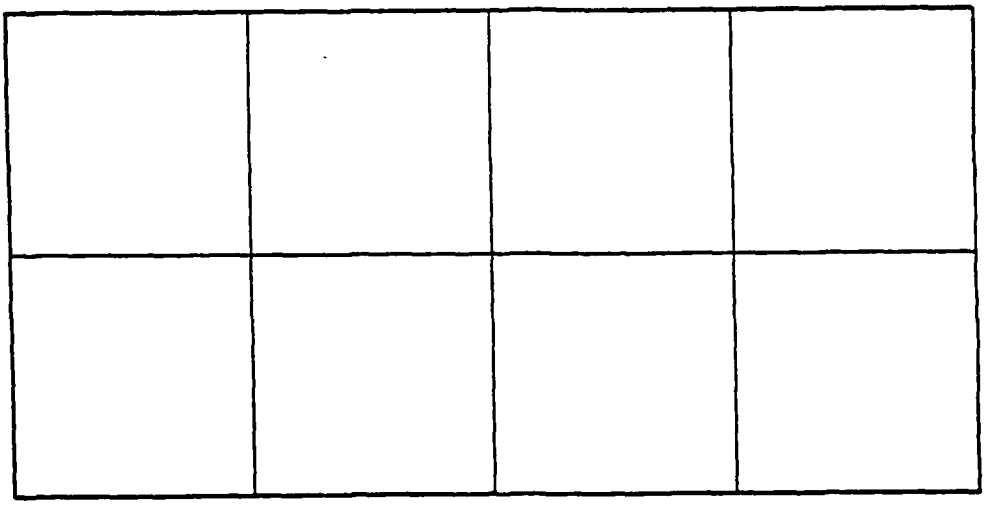
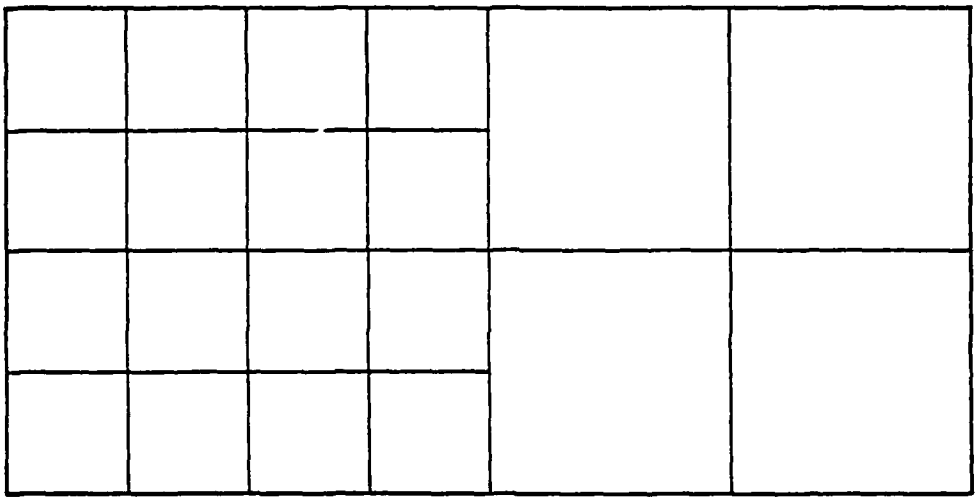

u = 0
v = 1
$X_n = \tau_{xy} = 0$
$Y_n = \sigma_y = 0$
u = 0
$Y_n = \tau_{xy} = 0$
$X_n = \sigma_x = 0$
v = 0
$X_n = \tau_{xy} = 0$
$Y_n = \sigma_y = 0$

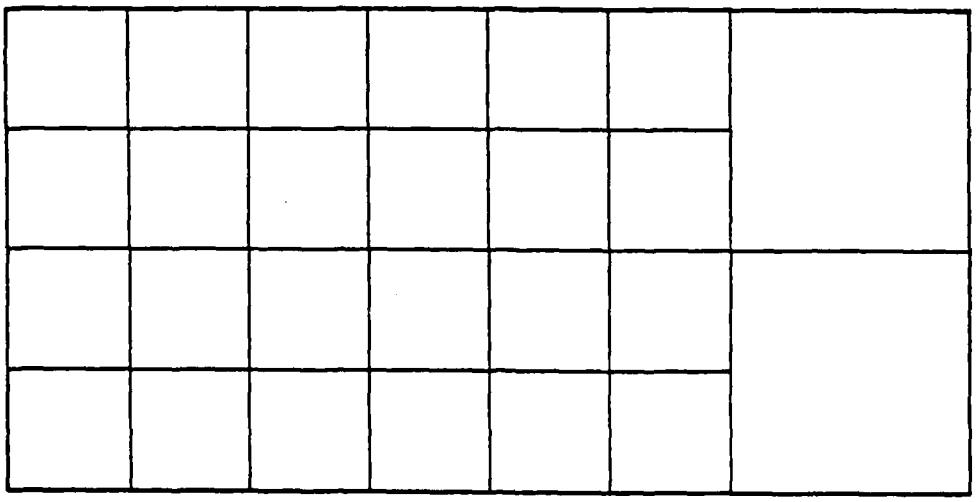
Figure 15



(a)

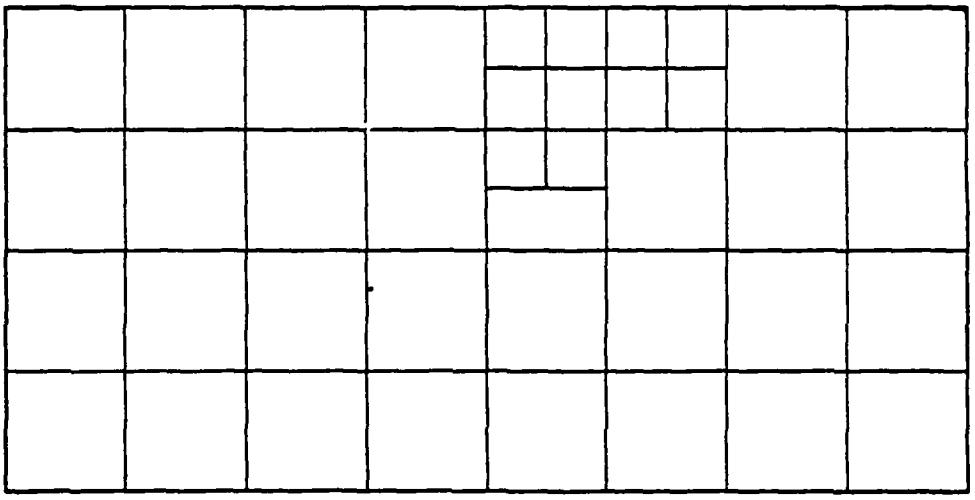


(b)

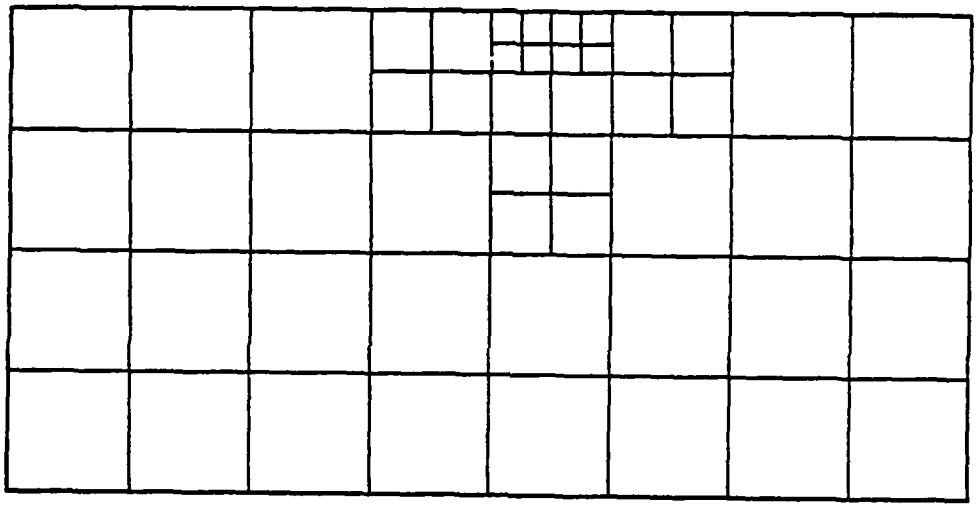


(c)

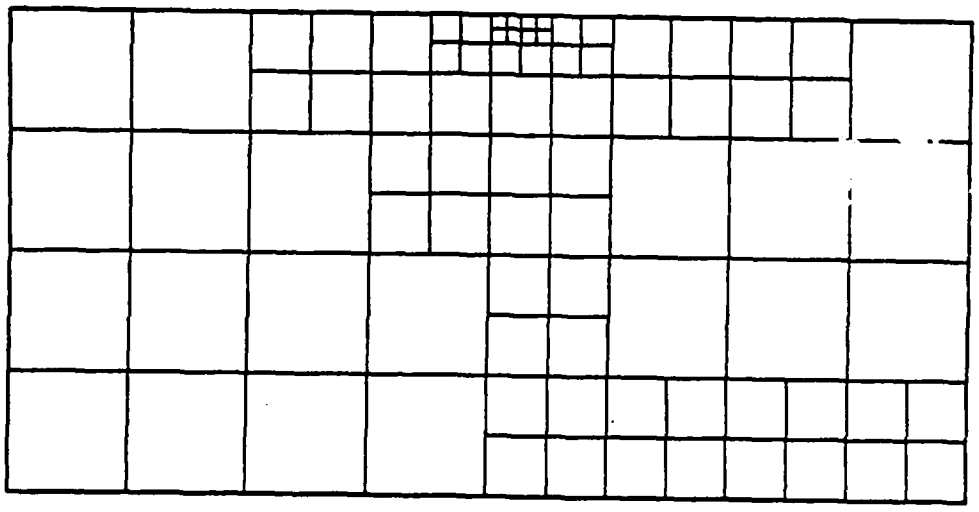
Figure 16



(d)

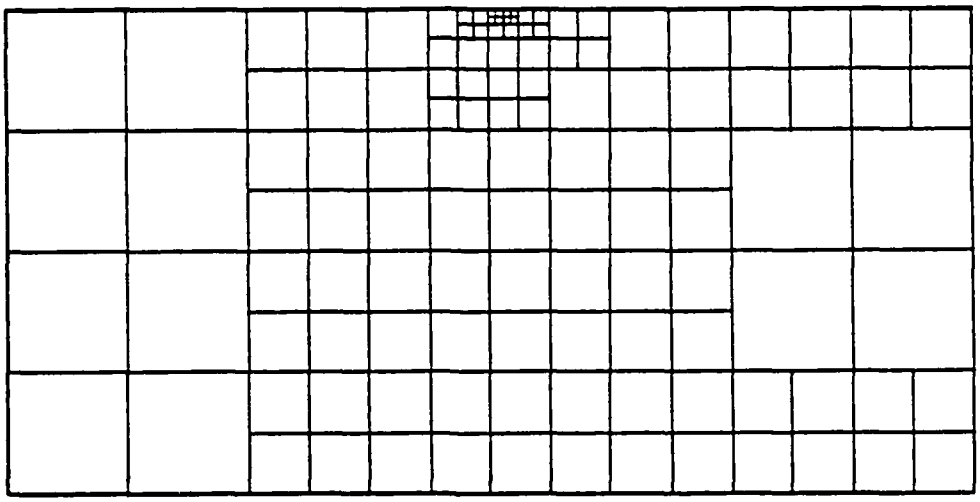


(e)

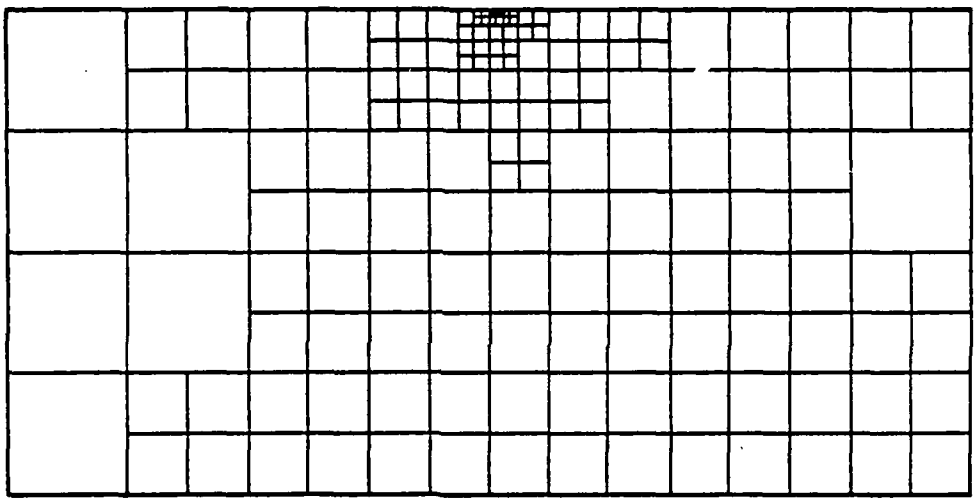


(f)

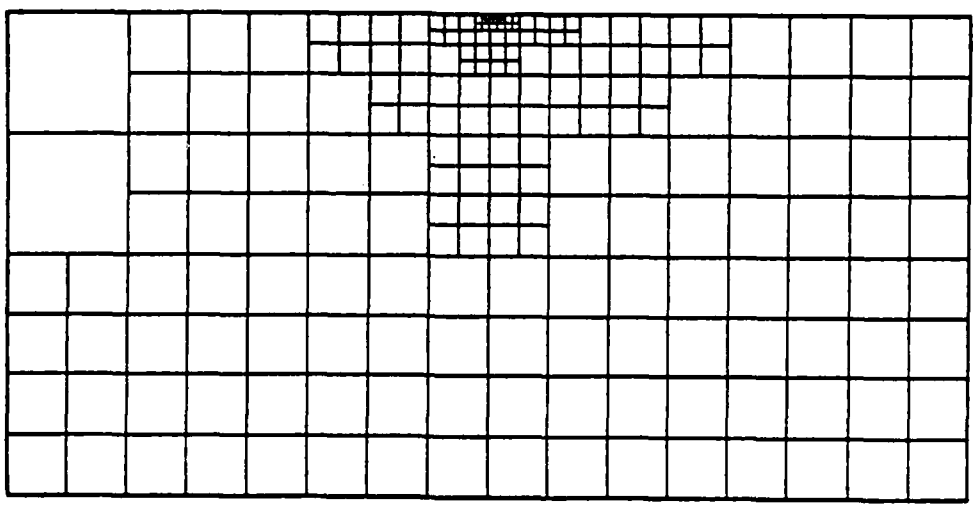
Figure 16



(g)

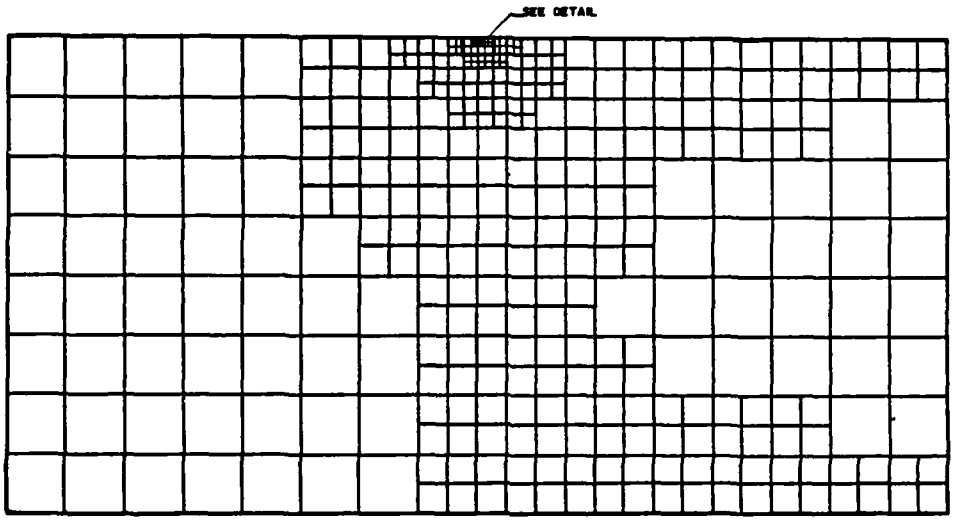


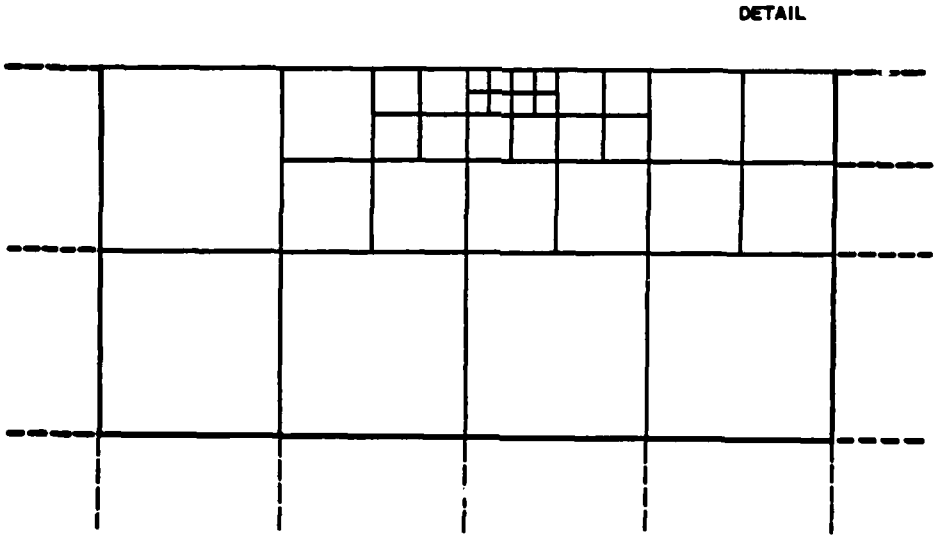
(h)



(i)

Figure 16





(j)

Figure 16

nine consecutive meshed generated by FEARS. The behavior of the (energy-norm) error for this problem in dependence of the degrees of freedom N is shown in Figure 17. Note that $N^{-1/2}$ is the fastest possible rate of convergence while the rate for uniform meshes is only $N^{-1/4}$ because of the presence of the singularity. Hence, we see from Figure 17 that the adaptive procedure provides us with a sequence of solutions with the fastest possible rate of convergence. Finally Figure 18 gives the effectivity index (1.1) for the estimator used in the process. Obviously, for errors of 10% or better the accuracy is very high.

Example 2: As a second elasticity problem we consider the behavior of isotropic, homogeneous material with $E = 1$, $\nu = 0$ on a ring-domain defined by concentric rings with radii 0.2 and 1, respectively. On the inner ring we specify the boundary conditions $u = 1$, $v = 0$ and on the outer ring $u = 0$, $v = 0$. Because of the symmetries of the solution, we restrict the computation to a quarter of the ring with obvious boundary conditions on the radial part of the boundary. As in Example 1 the error is measured in the energy norm. The sequence of meshes generated by FEARS is shown in Figure 19 and the corresponding error behavior is given in Table 13.

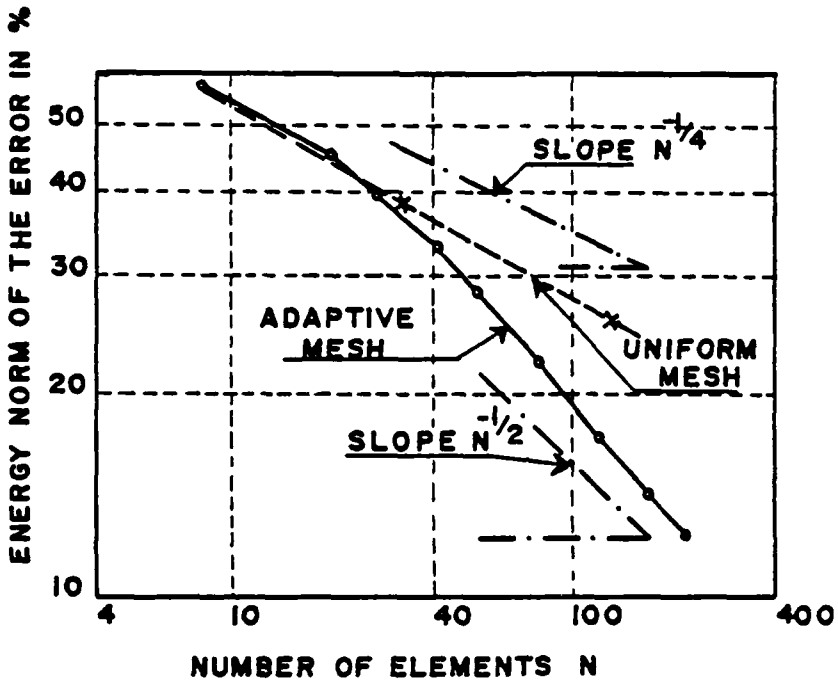

ENERGY NORM OF THE ERROR IN %
50
40
30
20
10
4
10
40
100
400
NUMBER OF ELEMENTS N
SLOPE N^-1/4
ADAPTIVE MESH
UNIFORM MESH
SLOPE N^-1/2

Figure 17

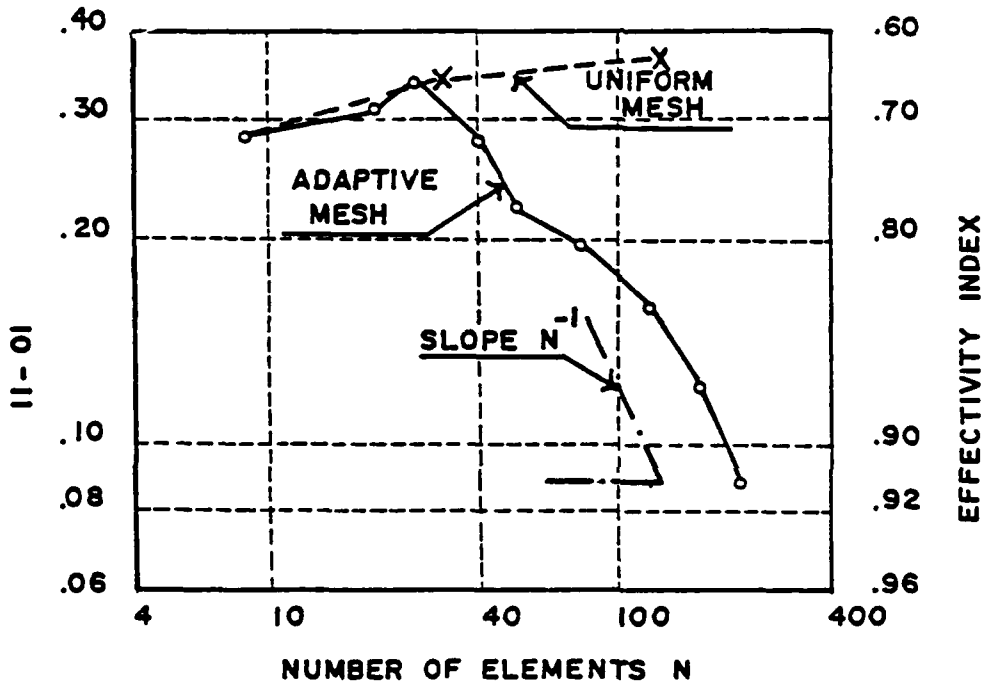

.40
.30
.20
.10
.08
.06
.60
.70
.80
.90
.92
.96
4
10
40
100
400
NUMBER OF ELEMENTS N
EFFECTIVITY INDEX
UNIFORM MESH
ADAPTIVE MESH
SLOPE N^-1

Figure 18

No. of Elem.	Degrees of Freedom	Actual Error $\|\|e\|\|_E$	Error Estim.	Rel. Error $100\,\|\|e\|\|/\|\|u\|\|$	Effect. θ
4	4	0.2495	0.2210	29.9%	0.885
16	24	0.1392	0.1135	16.7%	0.815
48	80	0.0784	0.0738	9.4%	0.942
145	225	0.0439	0.0434	5.3%	0.988

Table 13

Once again we see that the effectivity index θ of (1.1) has practically acceptable values when the accuracy of the solution is in the range of 5-10%.

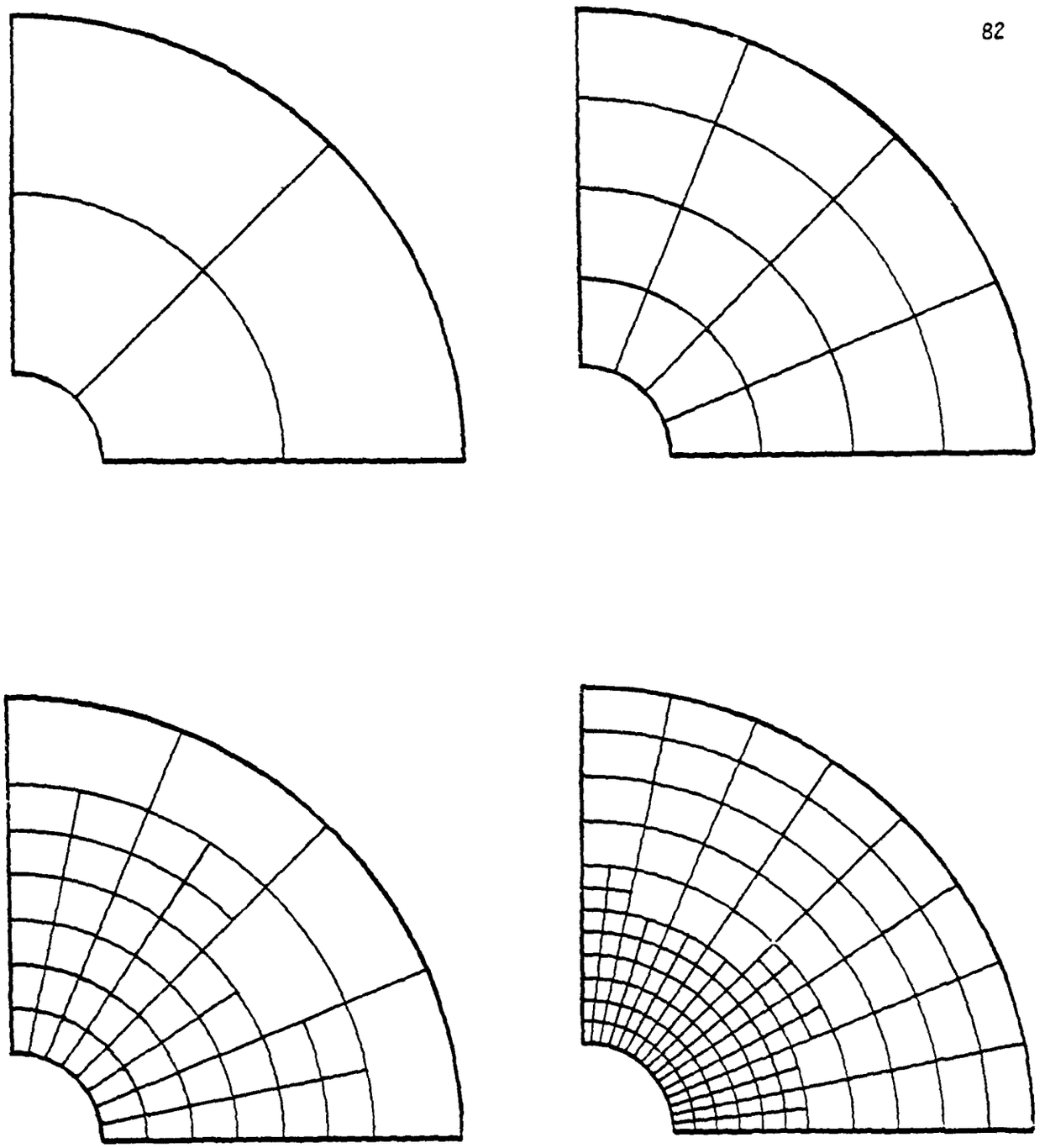


Figure 19

11. A Parabolic Problem

The ideas and approaches discussed here can also be applied to time-dependent parabolic problems. We restrict our comments to the simple model problem

$$\frac{\partial u}{\partial t} = a \frac{\partial^2 u}{\partial x^2}, \quad 0 < t < T, \quad 0 < x < 1 \tag{11.1}$$

with the initial conditions

$$u(0,x) = g(x) \tag{11.2}$$

and boundary conditions

$$u(t,0) = u(t,1) = 0. \tag{11.3}$$

The computations involved in the method of lines consist of the following two parts:

(11.4)
- (i) Use a finite-element discretization of the space variable to obtain an initial value problem for a system of ordinary differential equations.
- (ii) Solve this system numerically by means of some adaptive solver.

The discretizations in space and time have a different character. However, the objective is to equilibrate these two errors. We sketch here the principal idea of the error estimate of the finite element discretization and refer for more details to [3].

If the time-derivative du/dt on the left of (11.1) were given, then

the problem becomes elliptic and we know how to estimate the error in the energy norm

$$(11.5) \qquad ||e||_E^2 = \int_0^1 a\left(\frac{de}{dx}\right)^2 dx.$$

It is possible to show that with the values of du/dt defined by an approximate solution of the system of differential equations (11.4) (ii) we still obtain in this way a correct error estimator with an effectivity index that converges to one as the error tends to zero (see [3]).

We illustrate this approach on our model problem (11.1)-(11.3) with $g(x) = \sin \pi x$, $a = 0.3$ and $T = 1$ for which the differential equations (11.4) (ii) can be solved exactly. Table 14 shows the values of the effectivity index in the energy norm (11.5) at the time steps $t = 1/3, 2/3, 1$ in dependence of the number n of linear elements in a uniform mesh on the x-interval [0,1].

n \ t	1/3	2/3	1
4	0.911258	0.814769	0.713659
8	0.977357	0.948020	0.910430
16	0.994310	0.986583	0.976027
32	0.998575	0.996618	0.993897
64	0.999643	0.999152	0.999614
128	0.999909	0.999786	0.999892

Table 14

Evidently the effectivity index converges to one as n tends to infinity.

The above results assume that the differential equations (11.4) (ii) are solved explicitly. In practice, of course, this is not the case and we have to use one of the modern ODE-solvers as, for instance, LSODI which handles systems in the linearly implicit form

$$A(t,y)\dot{y} = b(t,y). \tag{11.6}$$

The solver selects adaptively the stepsize and order so as to meet a specified tolerance τ. Hence, our objective will be to choose the tolerance τ as a function of the space discretization and the step and order selection criterion of the ODE solver such that

(i) the error of the semidiscrete method is preserved at the least cost, and

(ii) the effectivity of the estimators is maintained.

These two objectives can be achieved with considerable success. We refer for details to [3].

References

[1] S. S. Antman, G. Rosenfeld, Global Behavior of Buckled States of Nonlinearly Elastic Rods, SIAM Review 20, 1978, 513-566. Correction: SIAM Review 22, 1980, 186-187.

[2] I. Babuška, M. R. Dorr, Error Estimates for the Combined H and P Versions of the Finite Element Method, Univ. of Maryland, Inst. f. Phys. Sci. and Techn., Tech. Note BN-951, 1980; to appear in Num. Math. 1981.

[3] I. Babuška, M. Luskin, An Adaptive Time Discretization Procedure for Parabolic Problems, in "Proc. of the Fourth IMACS Int. Symp. June 30 - July 2, 1981" ed. by R. Vichnevetsky and R. S. Stepleman, IMACS 1981, 5-9.

[4] I. Babuška, A. Miller, A Posteriori Error Estimates and Adaptive Techniques for the Finite Element Method, Univ. of Maryland, Inst. f. Phys. Sci. and Techn., Tech. Note BN-968, 1981.

[5] I. Babuška, W. Rheinboldt, Error Estimates for Adaptive Finite Element Computations, SIAM J. Num. Analysis 15, 1978, 736-754.

[6] I. Babuška, W. Rheinboldt, A Posteriori Error Estimates for the Finite Element Method, Intern. Journal f. Num. Methods in Eng., 12, 1978, 1597-1615.

[7] I. Babuška, W. Rheinboldt, A Posteriori Bounds and Adaptive Procedures for the Finite Element Method, in "Recent Advances in Engin. Science" ed. by R. L. Sierakowski, Univ. of Florida Press, Gainesville, FL, 1978, 413-418.

[8] I. Babuška, W. Rheinboldt, On a System for Adaptive, Parallel Finite Element Computations, in "Proc. 1978 Annual Conference", Association for Computing Mach., New York, 1978, 480-489.

[9] I. Babuška, W. Rheinboldt, Analysis of Optimal Finite Element Meshes in R^1, Math. of Comp. 33, 1979, 435-463.

[10] I. Babuška, W. Rheinboldt, On the Reliability and Optimality of the Finite Element Method, Computers and Structures 10, 1979, 87-94.

[11] I. Babuška, W. Rheinboldt, Adaptive Approaches and Reliability Estimations in Finite Element Analysis, Comp. Meth. in Appl. Mech. and Eng. 17/18, 1979, 519-540.

[12] I. Babuška, W. Rheinboldt, Reliable Error Estimation and Mesh Adaptation for the Finite Element Method, in "Computational Methods in Nonlinear Mechanics" ed. by J. T. Oden, North Holland Publ. Co., Amsterdam, 1980, 67-108.

[13] I. Babuška, W. Rheinboldt, A Posteriori Error Analysis of Finite Element Solutions for One-Dimensional Problems, SIAM J. Num. Anal. 18, 1981, 565-589.

[14] I. Babuška, B. A. Szabo, On the rates of Convergence of the Finite Element Method, Washington Univ., Center for Comp. Mechanics, Report WU/CCM-80/2, 1980; to appear in Int'l J. of Num. Meth. in Eng., 1981.

[15] I. Babuška, B. A. Szabo, I. N. Katz, The P-Version of the Finite Element Method, SIAM J. Num. Anal. 18, 1981, 515-545.

[16] F. Brezzi, J. Rappaz, P. A. Raviart, Finite Dimensional Approximation of Nonlinear Problems, Part I: Branches of Nonsingular Solutions, Ecole Polytechn., Centre de Math. Appl., Tech. Report 52, 1979.

[17] F. Brezzi, J. Rappaz, P. A. Raviart, Finite Dimensional Approximation of Nonlinear Problems, Part II: Limit Points, Ecole Polytechn. Report 64, 1980.

[18] F. Brezzi, J. Rappaz, P. A. Raviart, Finite Dimensional Approximation of Nonlinear Problems, Part III: Simple Bifurcation Points, Ecole Polytechn., Centre de Math. Appl., Tech. Report 65, 1980.

[19] C. den Heijer and W. Rheinboldt, On Steplength Algorithms for a Class of Continuation Methods, Univ. of Pittsburgh, Inst. f. Comp. Math. and Appl., Tech. Report ICMA-80-16, 1980; to appear in SIAM J. Num. Anal., 1981.

[20] R. Gaines, Difference Equations Associated with Boundary Value Problems for Second Order Nonlinear Ordinary Differential Equations, SIAM J. Num. Anal. 11, 1974, 411-434.

[21] G. H. Gonnet, On the Structure of Zero Finders, BIT 17, 1977, 170-183.

[22] J. H. Holland, Adaptation in Natural and Artificial Systems, Univ. of Michigan Press, Ann Arbor, MI, 1975.

[23] H. B. Keller, Numerical Solution of Bifurcation and Nonlinear Eigenvalue Problems, in "Applications of Bifurcation Theory" ed. by P. Rabinowitz, Academic Press, New York, NY, 1977, 359-384.

[24] H. B. Keller, Global Homotopies and Newton Methods, in "Recent Advances in Numerical Analysis" ed. by C. deBoor, G. H. Golub, Academic Press, New York, NY, 1978, 73-94.

[25] A. D. Kerr, M. T. Soifer, The Linearization of the Prebuckling State and its Effect on the Determined Instability Load, Trans. ASME, J. of Appl. Mech., V. 36, 1969, 775-783.

[26] S. T. Mau, R. H. Gallagher, A Finite Element Procedure for Nonlinear Prebuckling and Initial Postbuckling Analysis, NASA Contractor Report NASA-CR-1936, January 1972.

[27] G. Moore, A. Spence, The Calculation of Turning Points of Nonlinear Equations, SIAM J. Num. Anal. 17, 1980, 567-576.

[28] H. O. Peitgen, D. Saupe, K. Schmitt, Nonlinear Elliptic Boundary Value Problems Versus Their Finite Difference Approximations: Numerically Irrelevant Solutions, J. reine angew. Math. 322, 1981, 74-117.

[29] G. Pönisch, H. Schwetlick, Computing Turning Points of Curves Implicity Defined by Nonlinear Equations Depending on a Parameter, Computing 26, 1981, 107-121.

[30] T. Poston, I. Stewart, Catastrophe Theory and its Applications, Pitman Publ. Ltd., London 1978.

[31] W. Rheinboldt, Numerical Methods for a Class of Finite Dimensional Bifurcation Problems, SIAM J. Num. Anal. 15, 1978, 1-11.

[32] W. Rheinboldt, Solution Fields of Nonlinear Equations and Continuation Methods, SIAM J. Num. Anal. 17, 1980, 221-237.

[33] W. Rheinboldt, Adaptive Mesh Refinement Processes for Finite Element Solutions, Int. J. f. Num. Meth. in Eng. 17, 1981, 649-662.

[34] W. Rheinboldt, Numerical Analysis of Continuation Methods for Nonlinear Structural Problems, Computers and Structures 13, 1981, 103-114.

[35] W. Rheinboldt, Computation of Critical Boundaries on Equilibrium Manifolds, Univ. of Pittsburgh, Inst. f. Comp. Math. and Appl., Tech. Report ICMA-80-20, 1980; to appear in SIAM J. Num. Anal., 1981.

[36] W. Rheinboldt, C. K. Mesztenyi, On a Data-Structure for Adaptive Finite Element Mesh Refinements, ACM Trans. on Math. Software 6, 1980, 166-187.

[37] M. J. Sewell, On the Connection Between Stability and the Shape of the Equilibrium Surface, J. Mech. Phys. Solids 14, 1966, 203-230.

[38] M. J. Sewell, Some global equilibrium surface, Int. J. Mech. Engng. Educ. 6, 1978, 163-174.

[39] R. Seydel, Numerical Computation of Branch Points in Nonlinear Equations, Numer. Math. 33, 1979, 339-352.

[40] R. B. Simpson, A Method for the Numerical Determination of Bifurcation States of Nonlinear Systems of Equations, SIAM J. Num. Anal. 12, 1975, 439-451.

[41] B. A. Szabo, Some Recent Developments in Finite Element Analysis, Comp. Math. Appl. 5, 1979, 99-115.

[42] Ya. Z. Tsypkin, Adaptation and Learning in Automatic Systems; (transl. from the Russian by Z. J. Nicolic), Academic Press, NY, 1971.

[43] A. C. Walker, A Nonlinear Finite Element Analysis of Shallow Circular Arches, Int. J. Solids Structures 5, 1969, 97-107.

[44] L. A. Zadeh, On the Definition of Adaptivity, Proc. I.R.E. 51, 1963, 3-4.

[45] P. Zave, G. E. Cole, Jr., A Quantitative Evaluation of the Feasibility of, and Suitable Hardware Architectures for, an Adaptive, Parallel Finite Element System, Univ. of Maryland, Dept. of Comp. Science, Tech. Report, 1981, (in preparation).

[46] P. Zave, W. Rheinboldt, Design of an Adaptive Parallel Finite Element System, ACM Transactions on Math. Software 5, 1979, 1-17.

END

DATE